

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 05:43:30 ; Search time 1815.27 Seconds
(without alignments) 3677.450 Million cell updates/sec

Title: US-09-824-286-5
Perfect score: 319

Sequence: 1 GATATCGTATGACCCAGTC.....AGGACCAAGCTGGAGATCT 319

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%
	Listing first 45 summaries

```
Database :
1:  GenEmbl.*
2:  gb_ba.*
3:  gb_hhg.*
4:  gb_in.*
5:  gb_om.*
6:  gb_ov.*
7:  gb_pat.*
8:  gb_pl.*
9:  gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_hhg_hum.*
31: em_hhg_inv.*
32: em_hgtg_inv.*
33: em_hgtg_inv.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				Description
No.	Score	Match	Length	ID	

1	291.4	91.7	324	10	MUSNCN1	M853338 Mus musculus
2	292.4	91.3	321	10	AF163743	AF163743 Mus musculus
3	289.8	90.8	393	10	AF004328	AF004328 Mus muscu
4	288.4	90.3	353	10	AY058908	AY058908 Mus muscu
5	283.8	89.0	324	10	MUS1GKX50	L16627 Mus musculus
6	281.4	88.8	402	10	MUS1GKX50	X05877 Mouse mrna
7	281.8	88.3	909	6	AR007979	AR007979 Sequence
8	281.8	88.3	909	6	AR058994	AR058994 Sequence
9	281.8	88.3	909	6	123444	123444 Sequence 1
10	281.8	88.3	909	6	113364	113364 Sequence 7
11	281.8	88.3	909	6	161419	161419 Sequence 7
12	281.8	88.3	909	6	196033	196033 Sequence 7
13	279.6	87.6	318	10	MUS1GMRN1	L21025 Mus musculus
14	278.6	87.3	318	10	MUS1GMAF	M6251 Mouse Ig ka
15	276.4	86.6	321	10	MUS1GKX25	M34120 Mouse Ig ac
16	274.8	86.1	324	10	MUS1GKX75	L16623 Mus musculus
17	273.8	85.8	321	10	MMNPCR1B6	X70097 M.musculus
18	272.2	85.3	321	10	MUS1GKAD	L20956 Mus musculus
19	272.2	85.3	324	10	MUS1GKX77	L16638 Mus musculus
20	272.2	85.3	456	10	MUS1GKXK	M19614 Mouse Ig re
21	271.8	85.2	324	10	MUS1GKX08	L16617 Mus musculus
22	271.6	85.1	321	6	132985	132985 Sequence 39
23	271.6	85.1	324	10	MUS1GKX57	L16632 Mus musculus
24	271.6	85.1	324	10	MUS1GKX59	L16634 Mus musculus
25	271.6	85.1	327	10	AF192257	AF192257 Mus muscu
26	271.6	85.1	360	6	132994	132994 Sequence 48
27	271.6	85.1	360	23	E09038	E09038 CDNA encodi
28	271.6	84.8	321	10	MMNPCR1B1	X70095 M.musculus
29	270.6	84.8	324	10	AF087027	AF087027 Mus muscu
30	270.6	84.8	381	10	AF001496	AF001496 Mus muscu
31	270.2	84.7	324	10	MUS1GKX11	L16615 Mus musculus
32	269.6	84.5	324	10	MUS1GKX31	L16625 Mus musculus
33	269.2	84.5	721	10	MMU235962	AJ235962 Mus muscu
34	269.2	84.4	324	10	MUS1GKX17B	L16621 Mus musculus
35	269	84.3	384	10	MUS1KCA	L14875 Mus musculus
36	268.8	84.3	313	10	MUS1GKXAF	M64165 Mus musculus
37	267.8	83.9	320	10	MUS1GKX6A	L19553 Mus musculus
38	267.4	83.8	381	10	AF045512	AF045512 Mus muscu
39	266.8	83.6	303	10	MUS1GALFE	M36250 Mouse Ig ka
40	266.8	83.6	321	10	MMALCVR23	X06988 M.musculus
41	266.8	83.6	362	10	AF276281	AF276281 Mus muscu
42	266.4	83.5	324	10	MUS1GKX56	L16630 Mus musculus
43	266	83.4	324	10	MUS1GKX63	L16636 Mus musculus
44	265.8	83.3	321	10	MUS1GHMB	L32698 Mouse reart
45	265.8	83.3	341	10	MUS1GKXBP	M1761 Mouse Ig ka

ALIGNMENTS

RESULT	1
MUSNC41	
LOCUS	
DEFINITION	Mus musculus NC41 mRNA, partial cds.
ACCESSION	M83538
VERSION	M83538.1
KEYWORDS	antibody light chain variable domain VL.
SOURCE	Mus musculus Spleen cDNA to mRNA.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 324)
AUTHORS	Hudson,P.J., Tulip,R., Varghese,J., Laver,W.G., Webster,R. and Colman,P.M.
TITLE	Refined Structure of the Influenza virus N9 Neuraminidase-NC41 Fab complex
JOURNAL	Unpublished (1991)
FEATURES	
source	Location/Qualifiers
	1..324
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/tissue_type="Spleen"
	1..324
gene	


```
CDS      1.336
         /note="WRI7 VL"
```

```

V_region      1. .285
              /gene="Igk-V19"
              /standard_name="V19"
gene          1. .324
              /gene="Igk-V19"
              70. .102
misc_feature  /gene="Igk-V19"
              /standard_name="CDRI1"

```

```
misc_feature 148..168
                /gene="Igk-V19"
                /standard_name="CDR2"
misc_feature 265..291
                /gene="Igk-V19"
                /standard_name="CDR3"
J-segment 286..324
                /gene="Igk-V19"
                /standard_name="J1"
BASE COUNT 82 a 81 c 79 g 74 t 8 others
ORIGIN
Query Match 89.0%; Score 283.8; DB 10; Length 324;
Best Local Similarity 92.5%; Pred. No. 2.6e-85;
Matches 294; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
OY 1 GATATCGTAATGACCCAGTCTCACAATTCATGTCACATGATGAGACAGTATCAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GACATTTGATGATGACCCAGTCTCACAATTCATGTCACATGATGAGACAGTATCAC 60
OY 61 ATCACCCTCAAGGCCAGTCAGATGTGACTACTGCTGAGCTGGTATCAACAAAACCA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ATCACCCTCAAGGCCAGTCAGATGTGACTACTGCTGAGCTGGTATCAACAAAACCA 120
OY 121 GGGCAATCTCTTAACCTTGATTTACTGCGCATCCACCGGCACACTGGAGTCCCTGAT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GGGCAATCTCTTAACCTTGATTTACTGCGCATCCACCGGCACACTGGAGTCCCTGAT 180
OY 181 CGCTTCACAGGAGTGATGATGTCGGACAGATTATCTCACCATTAGAGAGTGGAGGCT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CGCTTCACAGGAGTGATGATGTCGGACAGATTATCTCACCATTAGAGAGTGGAGGCT 240
OY 241 GAAGACCTGGCACTTTATTTACTGTCAGCAACATTATATCACTCCGTGACGTTGGTGA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 GAAGACCTGGCACTTTATTTACTGTCAGCAACATTATATCACTCCGTGACGTTGGTGA 300
OY 301 GGGACCAAGCTGGAGATC 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GGGACCAAGCTGGAGATC 318

RESULT 6
LOCUS MMIGKV 462 bp mRNA linear ROD 04-AUG-1992
DEFINITION Mouse mRNA for immunoglobulin kappa variable region.
ACCESSION X05877.1 GI:52195
VERSION X05877.1 GI:52195
KEYWORDS constant region; Ig kappa light chain; variable region.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 462)
AUTHORS Grant,F.J.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-1987) Dr. Francis Grant, Zymogenetics, Inc., 2121
North 35th Street, Seattle, WA 98103, USA
2 (bases 1 to 462)
AUTHORS Grant,F.J., Levin,S.D., Gilbert,T. and Kindsvogel,W.
TITLE Improved RNA sequencing method to determine immunoglobulin mRNA
sequence
JOURNAL Nucleic Acids Res. 15 (13), 5496 (1987)
MEDLINE 87260030
FEATURES
    source Location/Qualifiers
        1..462 /organism="Mus musculus"
        /strain="ML-05"
        /db_xref="taxon:10090"
        /cell_type="hybridoma"
        7..417 /note="variable region (AA 1-137)"
        CDS 7..>462
```

```
/codon_start=1
/product="V(kappa) gene product"
/protein_id="CA29301.1"
/db_xref="GI:52196"
/translation="MHRTSMGKIMKESIOIAFVFVFLMISGVDGIIVMOSHREKSTSV
GDRVSTICKASQDVTYAVSHYQOKPGOSPRLIFWASIRHRTGVDFRTGSSGDTYTL
TISVQADLDLALYCCQOHSTPLTFGAGTKLELRADAPVYSIFPSSSE"
misc_feature 418..>462
                /note="constant region (AA 138-152)"
BASE COUNT 114 a 121 c 112 g 115 t
ORIGIN
Query Match 88.8%; Score 283.4; DB 10; Length 462;
Best Local Similarity 93.4%; Pred. No. 3.6e-85;
Matches 296; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
OY 1 GATATCGTAATGACCCAGTCTCACAATTCATGTCACATGATGAGACAGTATCAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GACATTTGATGATGACCCAGTCTCACAATTCATGTCACATGATGAGACAGTATCAC 153
OY 61 ATCACCCTCAAGGCCAGTCAGATGTGACTACTGCTGAGCTGGTATCAACAAAACCA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 ATCACCCTCAAGGCCAGTCAGATGTGACTACTGCTGAGCTGGTATCAACAAAACCA 213
OY 121 GGGCAATCTCTTAACCTTGATTTACTGCGCATCCACCGGCACACTGGAGTCCCTGAT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GGGCAATCTCTTAACCTTGATTTACTGCGCATCCACCGGCACACTGGAGTCCCTGAT 180
OY 214 GGGCAATCTCTTAACCTTGATTTACTGCGCATCCACCGGCACACTGGAGTCCCTGAT 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 GGGCAATCTCTTAACCTTGATTTACTGCGCATCCACCGGCACACTGGAGTCCCTGAT 273
OY 181 CGCTTCACAGGAGTGATGATGTCGGACAGATTATCTCACCATTAGAGAGTGGAGGCT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 CGCTTCACAGGAGTGATGATGTCGGACAGATTATCTCACCATTAGAGAGTGGAGGCT 333
OY 241 GAAGACCTGGCACTTTATTTACTGTCAGCAACATTATATCACTCCGTGACGTTGGTGA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 GAAGACCTGGCACTTTATTTACTGTCAGCAACATTATATCACTCCGTGACGTTGGTGA 393
OY 301 GGGACCAAGCTGGAGAT 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 GGGACCAAGCTGGAGAT 410

RESULT 7
LOCUS AR007979 909 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 1 from patent US 5753204.
ACCESSION AR007979
VERSION AR007979.1 GI:3967088
KEYWORDS
SOURCE unknown.
ORGANISM unknown.
REFERENCE 1 (bases 1 to 909)
AUTHORS Huston,J.S., Houston,L.L., Ring,D.B. and Oppermann,H.
TITLE Biosynthetic binding proteins for immunotargeting
JOURNAL Patent: US 5753204-A 1 19-MAY-1998;
FEATURES
    source Location/Qualifiers
        1..909 /organism="unknown"
BASE COUNT 229 a 218 c 246 g 216 t
ORIGIN
Query Match 88.3%; Score 281.8; DB 6; Length 909;
Best Local Similarity 93.1%; Pred. No. 1.3e-84;
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
OY 1 GATATCGTAATGACCCAGTCTCACAATTCATGTCACATGATGAGACAGTATCAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 GATATCGTAATGACCCAGTCTCACAATTCATGTCACATGATGAGACAGTATCAC 467
OY 61 ATCACCCTCAAGGCCAGTCAGATGTGACTACTGCTGAGCTGGTATCAACAAAACCA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 468 ATCTCTGCAAGGCCAGTGCAGGATGTGATCTGCTAGCTGGTATCAACAAAACCA 527
Oy 121 GGGCAATCTCCTAAACTTCGATTTTACGGCATCCACCGGACACAGTGAGTCCCTGAT 180
Db 528 GGGCAATCTCCTAAACTTCGATTTTACGGCATCCACCGGACACAGTGAGTCCCTGAT 587
Oy 181 CGCTTCACAGGACAGTGTGAGACAGATTTACTCTCACCATTACAGCAGTGTGAGGCT 240
Db 588 CGCTTCACAGGACAGTGTGAGACAGATTTACTCTCACCATTACAGCAGTGTGAGGCT 647
Oy 241 GAAGACCTGGCAGCTTTTATTACTGTCCAGCAACATTATATCACTCCGTGAGCTTCGGTGA 300
Db 648 GAAGACCTGGCAGCTTTTATTACTGTCCAGCAACATTATATAGTCCGTACACGTTCCGAGGG 707
Oy 301 GGGACCAAGCTGGAGAT 317
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 8
AR058994 909 bp DNA linear PAT 29-SEP-1999
LOCUS AR058994
DEFINITION Sequence 1 from patent US 5837846.
ACCESSION AR058994 GI:5984571
VERSION AR058994.1
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 909)
AUTHORS Huston,J.S., Houston,L.L., Ring,D.B. and Oppermann,H.
TITLE Biosynthetic binding proteins for immuno-targeting
JOURNAL Patent: US 5837846-A 1 17-NOV-1998;
FEATURES
source 1..909
location/Qualifiers
BASE COUNT 229 a 218 c 246 g 216 t
ORIGIN

Query Match 88.3%; Score 281.8; DB 6; Length 909;
Best Local Similarity 93.1%; Pred. No. 1.3e-84;
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Oy 1 GATATCGTAAATGACCCAGTCTCACAATTCATGTCACATGAGTAGAGACGATACACC 60
Db 408 GATATCGTAAATGACCCAGTCTCACAATTCATGTCACATGAGTAGAGACGATACACC 467
Oy 61 ATCACTGCAAGGCCAGTGCAGATGTGACTGCTGAGCCTGGTATCAACAAAACCA 120
Db 468 ATCTCTGCAAGGCCAGTGCAGATGTGACTGCTGAGCCTGGTATCAACAAAACCA 527
Oy 121 GGGCAATCTCCTAAACTTCGATTTTACGGCATCCACCGGACACAGTGAGTCCCTGAT 180
Db 528 GGGCAATCTCCTAAACTTCGATTTTACGGCATCCACCGGACACAGTGAGTCCCTGAT 587
Oy 181 CGCTTCACAGGACAGTGTGAGACAGATTTACTCTCACCATTACAGCAGTGTGAGGCT 240
Db 588 CGCTTCACAGGACAGTGTGAGACAGATTTACTCTCACCATTACAGCAGTGTGAGGCT 647
Oy 241 GAAGACCTGGCAGCTTTTATTACTGTCCAGCAACATTATATCACTCCGTGAGCTTCGGTGA 300
Db 648 GAAGACCTGGCAGCTTTTATTACTGTCCAGCAACATTATATAGTCCGTACACGTTCCGAGGG 707
Oy 301 GGGACCAAGCTGGAGAT 317
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 9
LOCUS 123444 909 bp DNA linear PAT 07-OCT-1996
DEFINITION Sequence 1 from patent US 5534254.

ACCESSION I23444
VERSION I23444.1 GI:1603314
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 909)
AUTHORS Huston,J.S., Houston,L.L., Ring,D.B. and Oppermann,H.
TITLE Biosynthetic binding proteins for immuno-targeting
JOURNAL Patent: US 5534254-A 1 09-OCT-1996;
FEATURES
source 1..909
location/Qualifiers
BASE COUNT 229 a 218 c 246 g 216 t
ORIGIN

Query Match 88.3%; Score 281.8; DB 6; Length 909;
Best Local Similarity 93.1%; Pred. No. 1.3e-84;
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Oy 1 GATATCGTAAATGACCCAGTCTCACAATTCATGTCACATGAGTAGAGACGATACACC 60
Db 408 GATATCGTAAATGACCCAGTCTCACAATTCATGTCACATGAGTAGAGACGATACACC 467
Oy 61 ATCACTGCAAGGCCAGTGCAGATGTGACTGCTGAGCCTGGTATCAACAAAACCA 120
Db 468 ATCTCTGCAAGGCCAGTGCAGATGTGACTGCTGAGCCTGGTATCAACAAAACCA 527
Oy 121 GGGCAATCTCCTAAACTTCGATTTTACGGCATCCACCGGACACAGTGAGTCCCTGAT 180
Db 528 GGGCAATCTCCTAAACTTCGATTTTACGGCATCCACCGGACACAGTGAGTCCCTGAT 587
Oy 181 CGCTTCACAGGACAGTGTGAGACAGATTTACTCTCACCATTACAGCAGTGTGAGGCT 240
Db 588 CGCTTCACAGGACAGTGTGAGACAGATTTACTCTCACCATTACAGCAGTGTGAGGCT 647
Oy 241 GAAGACCTGGCAGCTTTTATTACTGTCCAGCAACATTATATCACTCCGTGAGCTTCGGTGA 300
Db 648 GAAGACCTGGCAGCTTTTATTACTGTCCAGCAACATTATATAGTCCGTACACGTTCCGAGGG 707
Oy 301 GGGACCAAGCTGGAGAT 317
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 10
LOCUS I43364 909 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 7 from patent US 5631158.
ACCESSION I43364
VERSION I43364.1 GI:2468608
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 909)
AUTHORS Dorel,H. and Oppermann,H.
TITLE Methods and compositions for high protein production from non-native DNA
JOURNAL Patent: US 5631158-A 7 20-MAY-1997;
FEATURES
source 1..909
location/Qualifiers
BASE COUNT 229 a 218 c 246 g 216 t
ORIGIN

Query Match 88.3%; Score 281.8; DB 6; Length 909;
Best Local Similarity 93.1%; Pred. No. 1.3e-84;
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Oy 1 GATATCGTAAATGACCCAGTCTCACAATTCATGTCACATGAGTAGAGACGATACACC 60

Db 408 GATATCGTATGACCCAGTCTCTTAATTCAATGTCACGTGAGGAGACGGGTGACG 467
QY 61 ATCAGCTCAAGGCGCAGTGTGACTGCTGTAGCTGTGATCAACAAAACCA 120
Db 468 ATCTCTGCAAGGCGCAGTGTGACTGCTGTAGCTGTGATCAACAAAACCA 527
QY 121 GGGCAATCTCTTAACCTTCTGATTTACTGGGCAATCCACCGGACAGTGGAGTCCCTGAT 180
Db 528 GGGCAATCTCTTAACCTTCTGATTTACTGGGCAATCCACCGGACAGTGGAGTCCCTGAT 587
QY 181 CGCTTCACAGGACAGTGTGATGTGGACATTTACTCTCACCATAGCAGTGTGAGGCT 240
Db 588 CGCTTCACAGGACAGTGTGATGTGGACATTTACTCTCACCATAGCAGTGTGAGGCT 647
QY 241 GAAGACCTGGCAGCTTATTACTGTGACAGCAATATATACCTCCGTGGACGTTGGTGA 300
Db 648 GAAGACCTGGCAGCTTATTACTGTGACAGCAATATATAGAGTCCGTACACGTTGGAGGG 707
QY 301 GGGACCAAGCTGGAGAT 317
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 11
LOCUS 161419
DEFINITION Sequence 7 from patent US 5658763.
ACCESSION 161419
VERSION 161419.1 GI:2479367
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 909)
AUTHORS Dorel, H. and Oppermann, H.
TITLE Methods and compositions for high protein production from non-native DNA
JOURNAL Patent: US 5658763-A 7 19-AUG-1997;
FEATURES
Location/Qualifiers
1..909
BASE COUNT 229 a 218 c 246 g 216 t
ORIGIN

Query Match 88.3%; Score 281.8; DB 6; Length 909;
Best Local Similarity 93.1%; Pred. No. 1.3e-84;
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 GATATCGTATGACCCAGTCTCTTAATTCAATGTCACGTGAGGAGACAGTATCAC 60
Db 408 GATATCGTATGACCCAGTCTCTTAATTCAATGTCACGTGAGGAGACAGTATCAC 467
QY 61 ATCAGCTCAAGGCGCAGTGTGACTGCTGTAGCTGTGATCAACAAAACCA 120
Db 468 ATCTCTGCAAGGCGCAGTGTGACTGCTGTAGCTGTGATCAACAAAACCA 527
QY 121 GGGCAATCTCTTAACCTTCTGATTTACTGGGCAATCCACCGGACAGTGGAGTCCCTGAT 180
Db 528 GGGCAATCTCTTAACCTTCTGATTTACTGGGCAATCCACCGGACAGTGGAGTCCCTGAT 587
QY 181 CGCTTCACAGGACAGTGTGATGTGGACATTTACTCTCACCATAGCAGTGTGAGGCT 240
Db 588 CGCTTCACAGGACAGTGTGATGTGGACATTTACTCTCACCATAGCAGTGTGAGGCT 647
QY 241 GAAGACCTGGCAGCTTATTACTGTGACAGCAATATATACCTCCGTGGACGTTGGTGA 300
Db 648 GAAGACCTGGCAGCTTATTACTGTGACAGCAATATATAGAGTCCGTACACGTTGGAGGG 707
QY 301 GGGACCAAGCTGGAGAT 317
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 12
LOCUS 196033
DEFINITION Sequence 7 from patent US 5733782.
ACCESSION 196033
VERSION 196033.1 GI:3940503
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 909)
AUTHORS Dorel, H. and Oppermann, H.
TITLE Methods and compositions for high protein production from non-native DNA
JOURNAL Patent: US 5733782-A 7 31-MAR-1998;
FEATURES
Location/Qualifiers
1..909
BASE COUNT 229 a 218 c 246 g 216 t
ORIGIN

Query Match 88.3%; Score 281.8; DB 6; Length 909;
Best Local Similarity 93.1%; Pred. No. 1.3e-84;
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 1 GATATCGTATGACCCAGTCTCTTAATTCAATGTCACGTGAGGAGACAGTATCAC 60
Db 408 GATATCGTATGACCCAGTCTCTTAATTCAATGTCACGTGAGGAGACAGTATCAC 467
QY 61 ATCAGCTCAAGGCGCAGTGTGACTGCTGTAGCTGTGATCAACAAAACCA 120
Db 468 ATCTCTGCAAGGCGCAGTGTGACTGCTGTAGCTGTGATCAACAAAACCA 527
QY 121 GGGCAATCTCTTAACCTTCTGATTTACTGGGCAATCCACCGGACAGTGGAGTCCCTGAT 180
Db 528 GGGCAATCTCTTAACCTTCTGATTTACTGGGCAATCCACCGGACAGTGGAGTCCCTGAT 587
QY 181 CGCTTCACAGGACAGTGTGATGTGGACATTTACTCTCACCATAGCAGTGTGAGGCT 240
Db 588 CGCTTCACAGGACAGTGTGATGTGGACATTTACTCTCACCATAGCAGTGTGAGGCT 647
QY 241 GAAGACCTGGCAGCTTATTACTGTGACAGCAATATATACCTCCGTGGACGTTGGTGA 300
Db 648 GAAGACCTGGCAGCTTATTACTGTGACAGCAATATATAGAGTCCGTACACGTTGGAGGG 707
QY 301 GGGACCAAGCTGGAGAT 317
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 13
LOCUS MUS1GMRNAJ
DEFINITION Mus musculus immunoglobulin light chain (IgL) mRNA, VJ region, partial cds.
ACCESSION L21025
VERSION L21025.1 GI:309391
KEYWORDS Ig light chain; V-region; V-region; immunoglobulin.
SOURCE Mus musculus (strain BALB/c, sub-species domesticus) male adult hyperimmunized spleen mRNA.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 318)
AUTHORS Iwasaki, Y., Takabatake, H. and Ferrone, S.
TITLE Anti-idiotypic monoclonal antibody variable region gene
JOURNAL Unpublished (1993)
FEATURES
Location/Qualifiers
1..318
/organism="Mus musculus"

```

/strain="BALB/c"
/sub_species="domesticus"
/db_xref="taxon:10090"
/map="chromosome 12"
/sex="male"
/cell_line="S2B154"
/cell_type="hybridoma"
/tissue_type="hyperimmunized spleen"
/dev_stage="adult"
1..285
/gene="Igh"
/exon
1..285
/gene="Igh"
/number=2
1..285
/gene="Igh"
<1..>318
/gene="Igl"
/codon_start=1
/product="Immunoglobulin light chain"
/protein_id="AA51170.1"
/db_xref="GI:309392"
/translation="DIVTQSHKPMSTSVGDRVSTICKASODVSTAVAWYQQRFGOSP
ILIIYSASRYRTGVDRFTSGSGSDFTFTISSVQAEIDLAVYCOQHITPWTGGGT
KLEI"
1..318
/gene="Igl"
286..318
/gene="Igl"
/number=3
286..318
/gene="Igl"
J-segment
BASE COUNT 81 a 80 c 77 g 80 t
ORIGIN

```

```

Query Match      87.6%; Score 279.6; DB 10; Length 318;
Best Local Similarity 92.5%; Pred. No. 7e-84;
Matches 294; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 GATATCGTAATGACCCAGTCTCACAATTCATGTCACATCAGTAGAGACAGTATCAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GACATTGTGATGACCCAGTCTCACAATTCATGTCACATCAGTAGAGACAGGTCAGC 60

QY 61 ATCACCCTGAAGGCGAGTGCAGATGTGACTGCTGTAGCCTGTATCAACAAAACCA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ATCACCCTGAAGGCGAGTGCAGATGTGACTGCTGTAGCCTGTATCAACAAAACCA 120

QY 121 GGGCAATCTCTAACTTCTGATTTACTGGGCAATCCAGCGACACTGAGTCCCTGAT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GGGCAATCTCTAACTTCTGATTTACTGGGCAATCCAGCGACACTGAGTCCCTGAT 180

QY 181 CGCTTCACAGCAGTGCAGATGTGAGATATATCTTCACCATAGCAGTGTGAGGCT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CGCTTCACAGCAGTGCAGATGTGAGATATATCTTCACCATAGCAGTGTGAGGCT 240

QY 241 GAAGACCTGGGACCTTATTTACTGTCAGCAACATTTATTCACCTGGAGCGTTGGTGA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 GAAGACCTGGGACCTTATTTACTGTCAGCAACATTTATTTACTCCGTGGACGTTGGTGA 300

QY 301 GGGACCAAGCTGGAGATC 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GGGACCAAGCTGGAAATC 318

```

```

RESULT 14
MUSIGLAF 318 bp mRNA linear ROD 27-APR-1993
LOCUS
DEFINITION Mouse Ig kappa-chain mRNA V region, partial cds, from hybridoma
H250-6.
M36251
ACCESSION M36251.1 GI:197641
VERSION
KEYWORDS V-region; Immunoglobulin kappa-chain; Immunoglobulin light chain;

```

```

SOURCE
ORGANISM
    processed gene.
    Mouse (strain BALB/c), CDNA to mRNA, from hybridoma H250-6.
    Mus musculus
REFERENCE
AUTHORS
    1 (bases 1 to 318)
TITLE
    Kavalier, J., Caton, A.J., Staudt, L.M., Schwartz, D. and Gerhard, W.
    A set of closely related antibodies dominates the primary antibody
    response to the antigenic site CB of the A/PR/8/34 Influenza virus
    hemagglutinin
JOURNAL
    J. Immunol. 145, 2312-2321 (1990)
MEDLINE
    90375932
COMMENT
    Draft entry and computer-readable sequence for [J. Immunol. (1990)
    In press] kindly submitted
    by J.Kavalier, 06-JUL-1990.
FEATURES
    source
    1..318
    /organism="Mus musculus"
    /strain="BALB/c"
    /sub_species="domesticus"
    /db_xref="taxon:10090"
    /map="chromosome 6"
    /cell_line="H250-6"
    /tissue_type="hybridoma"
    1..318
    /gene="Igm"
    <1..>318
    /gene="Igm"
    /codon_start=1
    /product="Immunoglobulin kappa-chain VK-1"
    /protein_id="AA539110.1"
    /db_xref="GI:197642"
    /translation="DIVTQSHKPMSTSVGDRVSTICKASODVSTAVAWYQQRFGOSP
    KLIISASRYRTGVDRFTSGSGSDFTFTISSVQAEIDLAVYCOQHITPWTGGGT
    KLEI"

```

```

BASE COUNT 83 a 81 c 77 g 76 t 1 others
ORIGIN

Query Match      87.3%; Score 278.6; DB 10; Length 318;
Best Local Similarity 92.1%; Pred. No. 1.5e-83;
Matches 293; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 GATATCGTAATGACCCAGTCTCACAATTCATGTCACATCAGTAGAGACAGTATCAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GACATTGTGATGACCCAGTCTCACAATTCATGTCACATCAGTAGAGACAGGTCAGC 60

QY 61 ATCACCCTGAAGGCGAGTGCAGATGTGACTGCTGTAGCCTGTATCAACAAAACCA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ATCACCCTGAAGGCGAGTGCAGATGTGACTGCTGTAGCCTGTATCAACAAAACCA 120

QY 121 GGGCAATCTCTAACTTCTGATTTACTGGGCAATCCAGCGACACTGAGTCCCTGAT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GGGCAATCTCTAACTTCTGATTTACTGGGCAATCCAGCGACACTGAGTCCCTGAT 180

QY 181 CGCTTCACAGCAGTGCAGATGTGAGATATATCTTCACCATAGCAGTGTGAGGCT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CGCTTCACAGCAGTGCAGATGTGAGATATATCTTCACCATAGCAGTGTGAGGCT 240

QY 241 GAAGACCTGGGACCTTATTTACTGTCAGCAACATTTATTCACCTGGAGCGTTGGTGA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 GAAGACCTGGGACCTTATTTACTGTCAGCAACATTTATTTACTCCGTGGACGTTGGTGA 300

QY 301 GGGACCAAGCTGGAGATC 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GGGACCAAGCTGGAAATC 318

```

```

RESULT 15
MUSIGCSR 321 bp mRNA linear ROD 27-APR-1993
LOCUS
DEFINITION Mouse Ig active kappa-chain mRNA V-region from hybridoma GA733,
partial cds.

```

ACCESSION M34120
 VERSION M34120.1 GI:197289
 KEYWORDS Immunoglobulin light chain; Immunoglobulin-kappa; processed gene.
 SOURCE Mouse (Balb/c) hybridoma GA733, cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 321)
 AUTHORS Caton, A.J., Herlyn, D., Ross, A.H. and Koprowski, H.
 TITLE Identical D region sequences expressed by murine monoclonal antibodies specific for a human tumor-associated antigen
 JOURNAL J. Immunol. 144, 1965-1968 (1990)
 MEDLINE 90171596
 FEATURES
 location/Qualifiers
 1..321
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 <1..>321
 /note="Ig kappa-chain"
 /codon_start=1
 /protein_id="AAA39001.1"
 /db_xref="GI:197290"
 /translation="DIVMTOSHKFMSTSVGDSVITCKASQDSTAVAWYQKPGQSP
 KLIVSASDRYTGVPDRFTSGSGTDFFTISSVQAEIDLAVYCHQHYITPRTFGGT
 KLEIK"
 CDS
 BASE COUNT 87 a 83 c 74 g 77 t
 ORIGIN Chromosome 6.
 Query Match 86.6%; Score 276.4; DB 10; Length 321;
 Best Local Similarity 91.8%; Pred. No. 8.6e-83;
 Matches 292; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
 QY 1 GATATCGTAATGACCCAGTCTCACAATTCATGTCACATCAGTAGAGACAGATACACC 60
 Db 1 GACATTGTGATGACCCAGTCTCACAATTCATGTCACATCAGTAGAGACAGTGTGAGC 60
 QY 61 ATCACCCTGCAAGGCCAGTCAAGATGTGACTGCTGTAAGCCTGGTATCAACAAAACCA 120
 Db 61 ATCACCCTGCAAGGCCAGTCAAGATGTGACTGCTGTAAGCCTGGTATCAACAAAACCA 120
 QY 121 GGGCAATCTCCTAAACTCTGATTACTGCGGATCCACCCGACACATGAGTCCCTGAT 180
 Db 121 GGACAATCTCCTAAACTCTGATTACTGCGGATCCACCCGATCAGTACAGTCCCTGAT 180
 QY 181 CGCTTCACAGCAGTGTGATGCGACAGATTACTTCACCATCAGCAGTGTGACAGGCT 240
 Db 181 CGCTTCAGTGTGATGCGACAGATTACTTCACCATCAGCAGTGTGACAGGCT 240
 QY 241 GAAGACCTGGCACTTTATTAGTCAGCAACATTATATCCTCGTGGACGTTGGTGA 300
 Db 241 GAAGACCTGGCACTTTATTAGTCAGCAACATTATATCCTCGTGGACGTTGGTGA 300
 QY 301 GGGACCAAGCTGAGATC 318
 Db 301 GGCACCAAGCTGAGATC 318

Search completed: August 20, 2002, 09:25:29
 Job time: 13319 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 06:41:25 ; Search time 363.1 Seconds
(without alignments)
1508.388 Million cell updates/sec

Title: US-09-824-286-5
Perfect score: 319
Sequence: 1 GATATCGTATGACCCAGTC.....AGGACCAAGCTGAGATCT 319

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_032802.*
1: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	315.8	99.0	319	AA97440	Monoclonal antibody
2	281.8	88.3	909	AA736878	741F8 anti-c-erbB-
3	281.8	88.3	909	AA791837	Anti-c-erbB-2 sing
4	281.8	88.3	909	AA778879	Single chain antib
5	281.8	88.3	909	AA721796	741F8 anti-c-erbB-
6	281.8	88.3	909	AA722598	DNA encoding singl
7	281.8	88.3	909	AA763397	Anti-c-erbB-2 sFv
8	273.8	85.8	2001	AA046088	Sequence encoding
9	271.6	85.1	360	AA090434	DNA encoding anti-

10	268.4	84.1	321	AA090424	DNA encoding Immun
11	265.8	83.3	394	AAV93345	Mouse anti-HM1.24
12	265.8	83.3	394	AAV36286	DNA encoding L ch
13	261.2	81.9	321	AA090423	DNA encoding Immun
14	259.8	81.4	303	ABA94232	Mouse 04 antibody
15	259.6	81.4	420	AA090433	DNA encoding anti-
16	257.2	80.6	705	AA97063	DNA encoding TRA-8
17	256.2	80.3	380	AAA29700	CC92 VL region nuc
18	256.2	80.3	380	AA240721	Antibody C92 high
19	256.2	80.3	380	AA503550	Mouse DNA encoding
20	256.2	80.3	381	AA04265	Colon Cancer monoc
21	254	79.6	411	AA73901	Variable light sub
22	254	79.6	411	AAV38409	Antibody 24-31 hum
23	253	79.3	882	AA040022	cDNA insert of PK1
24	251.4	78.8	321	AA065552	Mouse anti-HIV m3
25	251.4	78.8	321	AA070370	Anti HIV antibody
26	251.4	78.8	882	AA040023	mRNA encoding Kapp
27	250.8	78.6	411	AA005557	Sequence encoding
28	250.6	78.6	321	AA025487	Human mAb 12B1 lig
29	249.8	78.3	321	AA294887	Tie2 receptor antia
30	248.6	77.9	748	AA05781	Erbb-2-specific sc
31	248.2	77.8	321	AA758265	Lead binding Mab 2
32	248.2	77.8	321	AAV81046	Murine 708 VL DNA
33	248.2	77.8	324	AA090813	DNA encoding light
34	248.2	77.8	324	AA004592	Sequence encoding
35	248.2	77.8	387	AA027132	Light chain variab
36	248.2	77.8	387	AA099034	Murine antibody Fv
37	248.2	77.8	407	AA028254	VL FRP5. Mus musc
38	247.6	77.6	382	AA094484	Plasmid pUC-M21-V
39	247.6	77.6	382	AA738599	Chimeric human/mu
40	247.6	77.6	794	AA742034	Plasmid pMW152-5 f
41	247.6	77.6	1371	AA250584	M79scFv-interleuk1
42	247.6	77.6	1389	AA250585	M79scFv-interleuk1
43	247.6	77.6	1479	AA705783	Leader-scFv(FRP5):
44	247.6	77.6	1692	AA729409	scFv(FRP5)-DETA-DG
45	247.6	77.6	1862	AA729413	DGAL4-DP-scFv(FRP5

ALIGNMENTS

RESULT 1	AA97440	standard; cDNA; 319 BP.
XX	AA97440:	
AC	AA97440:	
XX		
DT	21-MAY-1998	(first entry)
XX		
DE	Monoclonal antibody CP.B8 light chain variable region cDNA.	
XX		
KW	Cytokine receptor; gamma common chain; gc chain; human;	
KW	blocking agent; monoclonal antibody; CP.B8; Immunological disease;;	
KW	myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;	
KW	insulin-dependent diabetes; inflammatory bowel disease;	
KW	sympathetic ophthalmia; uveitis; allergy; asthma; infection;;	
KW	graft versus host disease; psoriasis; immunosuppressive; therapy;	
SS.		
XX		
OS	Mus musculus.	
XX		
PN	MO9743416-A1.	
XX		
PD	20-NOV-1997.	
XX		
PF	09-MAY-1997;	97WO-US07870.
XX		
PR	10-MAY-1996;	96US-0017466.
XX		
PA	(BIOJ) BIOGEN INC.	
XX		
PI	Benjamin CD, Burkly LC, Hession C, Whitty A;	

DR WPI: 1998-008885/01.
 DR P-PSDB: AAW31647.
 XX
 XX Blocking agents of the gamma common chain of cytokine receptors -
 PT particularly monoclonal antibodies, used to induce T cell anergy for
 PT treatment of immunological diseases
 PS
 XX Claim 21: Page 81-82; 11pp: English.
 PS
 XX This cDNA sequence codes for the light chain variable region (see
 CC AAW31647) of monoclonal antibody (Mab) C9.B8, which is produced
 CC by a hybridoma deposited as ATCC 12107, and which is specific for
 CC the gamma constant (gc) chain (see AAW31646) of human cytokine
 CC receptors. A Mab having complementarity determining regions
 CC encoded by the VL or VH (see AAT07441) cDNA is claimed. The
 CC invention provides compositions and methods for inhibiting cytokine
 CC signaling using gc chain blocking agents for the treatment of
 CC immunological diseases such as myasthenia gravis, rheumatoid
 CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,
 CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,
 CC allergy, asthma, parasitic infection, graft vs. host disease or
 CC psoriasis. Nucleic acids encoding the gc-blocking agents can be
 CC used in gene therapy for the same purposes. A preferred gc
 CC blocking agent in Mab C9.B8 or its Fab fragment.
 XX
 XX Sequence 319 BP; 86 A; 87 C; 71 G; 75 T; 0 other;
 SQ

Query Match 99.0%; Score 315.8; DB 19; Length 319;
 Best Local Similarity 99.4%; Pred. No. 1.4e-93;
 Matches 317; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GATATCGTAATGACCCAGCTTCACAATTATCATGTCACATGAGAGACAGTATCAC 60
 Db 1 gatacgtaatatgccagcttcacaattcatcacaatcagtagagacagatcaccc 60
 OY 61 ATCACTGCAAGGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 120
 Db 61 atcacctgcaagcgagtcagagtcagagtcagagtcagagtcagagtcagagtcag 120
 OY 121 GGGCAATCTCCTAACTTCTGATTACTGAGGATCCACCGGACACTGAGTCCCTGAT 180
 Db 121 gggcaatctcctaaacttctgattactgagtcagagtcagagtcagagtcagagtc 180
 OY 181 CGCTTCACAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 Db 181 cgcttcacagagagtcagagtcagagtcagagtcagagtcagagtcagagtcag 240
 OY 241 GAAGACCTGGCACTTATTACTGTCAGCAACATTATATCACTCCGTGAGCTGATGA 300
 Db 241 gaagacctggcacttattactgtcagcaaacattatatactccgtgagctgagtg 300
 OY 301 GGGACCAAGCTGAGATCT 319
 Db 301 gggaccaagctgagatct 319

RESULT 2
 AAT36878
 ID AAT36878 standard; cDNA: 909 BP.
 XX
 AC AAT36878;
 XX
 DT 29-OCT-1996 (first entry)
 XX
 DE 741F8 anti-C-erbB-2 two single chain Fv construct.
 XX
 KW 741F8: anti-C-erbB-2; monoclonal antibody; single chain Fv; sfv;
 KW construct; polypeptide linker; C-terminal amino acid sequence;
 KW in vivo imaging; drug targeting experiment; homodimer;
 KW increased; binding avidity; tissue retention time; ss.
 XX
 OS Homo sapiens.

XX
 FH Key Location/Qualifiers
 FT CDS 3..755
 FT CDS /*tag= a
 FT
 XX
 XX US5534254-A.
 PN
 XX
 PD 09-JUL-1996.
 XX
 XX 06-FEB-1992; 92US-0831967.
 PF
 XX 07-OCT-1993; 93US-0133804.
 PR
 XX 06-FEB-1992; 92US-0831967.
 PA
 XX (CHIR) CHIRON CORP.
 PA (CREA-) CREATIVE BIOMOLECULES INC.
 PI Houston LL, Huston JS, Oppermann H, Ring DB;
 XX
 XX WPI: 1996-333194/33.
 DR P-PSDB: AAW02278.
 XX
 XX Compns. confg. antigen-targeting antibody fragment constructs -
 PT comprising dimer of single-chain Fv fragments
 PT
 XX Example 1: Columns 27-28; 30pp: English.
 PS
 XX The variable heavy (VH) and variable light (VL) genes of the 741F8
 CC anti-C-erbB-2 monoclonal antibody (Mab), were isolated from the
 CC cDNA of the parental 741F8 hybridoma line. A two single chain Fv
 CC (sfv) gene was constructed by connecting the VH and VL genes with a
 CC DNA sequence encoding a polypeptide linker. A synthetic DNA duplex
 CC encoding the C-terminal amino acid sequence, (gly)4-Cys was
 CC inserted, and the resulting 741F8 anti-C-erbB-2 two sfv inserted
 CC into an expression vector. The resulting gene, the present
 CC sequence, was transformed into E. coli, and protein expression
 CC induced by the addn. of IPTG to the culture medium.
 CC A compsn. comprising a carrier and the 2 sfv protein prod. can be
 CC used for in vivo imaging, and drug targeting experiments. The
 CC 2 sfv protein prod. is a homodimer, in which both fragments target
 CC the same antigen, therefore giving greater binding avidity and
 CC longer tissue retention times, compared to individual sfv protein
 CC prod. fragments.
 XX
 SQ Sequence 909 BP; 229 A; 218 C; 246 G; 216 T; 0 other;

Query Match 88.3%; Score 281.8; DB 17; Length 909;
 Best Local Similarity 93.1%; Pred. No. 3e-82;
 Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 1 GATATCGTAATGACCCAGCTTCACAATTATCATGTCACATGAGAGACAGTATCAC 60
 Db 408 gatacgtaatatgccagcttcacaattcatcacaatcagtagagacagagtcagc 467
 OY 61 ATCACTGCAAGGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 120
 Db 61 atcacctgcaagcgagtcagagtcagagtcagagtcagagtcagagtcagagtcag 120
 OY 121 GGGCAATCTCCTAACTTCTGATTACTGAGGATCCACCGGACACTGAGTCCCTGAT 180
 Db 121 gggcaatctcctaaacttctgattactgagtcagagtcagagtcagagtcagagtc 180
 OY 181 CGCTTCACAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 Db 181 cgcttcacagagagtcagagtcagagtcagagtcagagtcagagtcagagtcag 240
 OY 241 GAAGACCTGGCACTTATTACTGTCAGCAACATTATATCACTCCGTGAGCTGATGA 300
 Db 241 gaagacctggcacttattactgtcagcaaacattatagagtcgtagacgctcgagtg 300
 OY 301 GGGACCAAGCTGAGATCT 317
 Db 301 gggaccaagctgagatct 317

Db 708 gggaccaagctgagat 724

RESULT 3

AAT91837

ID AAT91837 standard; cDNA; 909 BP.

XX

AC AAT91837;

XX

DT 07-JAN-1998 (first entry)

XX

DE Anti-c-erbB-2 single chain antibody 741F8 scfv gene.

XX

DE Adenovirus: E1A; transactivator; transcription activator; stimulate;

XX

KM expression vector; single-chain binding protein; VAI; enhance; PCR;

KM translation; production; immortal; eukaryotic cell; scfv; primer;

KM single-chain antibody fragment; imaging; tumour; breast cancer;

KM ovarian cancer; c-erbB-2 antigen; digoxin intoxication; ss.

XX

OS Synthetic.

XX

XX Key Location/Qualifiers

FT CDS 2..755

FT /*tag= a

XX

PN US5658763-A.

XX

PD 19-AUG-1997.

XX

PE 25-OCT-1993; 93US-0143498.

XX

PR 25-OCT-1993; 93US-0143498.

PR 05-JUN-1995; 95US-0463675.

XX

PA (CREA-) CREATIVE BIOMOLECULES INC.

PI Dorai H, Oppermann H;

PI WPI: 1997-424235/39.

DR P-PSDB; AAW29261.

XX

PT Producing single chain binding protein in immortalised eukaryotic

PT cells - which comprise protein coding sequences, a transcription

PT activator and translation promotion sequences, provides high

PT expression at low copy number

XX

PS Example 2; Column 29-32; 24pp; English.

XX

CC This sequence encodes a single chain anti-c-erbB-2 antibody (scfv)

CC derived from hybridoma 741F8. The protein product includes a C-terminal

CC tail Gly4-Cys in some constructs. Single-chain antibodies can be

CC produced using a novel method which comprises culturing an immortalised

CC eukaryotic cell having transfected DNA sequences (encoding the protein

CC of interest), integrated into its genome. In particular expression

CC effector vectors containing a non-native reporter DNA (encoding the scfv)

CC and viral sequences to promote transcription and translation (e.g. the

CC adenovirus E1A and VAI genes as shown in AAT91831 and AAT91834

CC respectively) are used. The scfv that is produced, when properly folded,

CC has a structure with mono- or bi-functional binding activity. The method

CC is especially used to produce single-chain antibody fragments (scfv),

CC e.g. for imaging tumours or delivering therapeutic agents to them,

CC particularly breast and ovarian cancers that express the c-erbB-2

CC antigen. Other scfv are used in model studies and for treating digoxin

CC intoxication.

XX

SO Sequence 909 BP; 229 A; 218 C; 246 G; 216 T; 0 other.

Query Match 88.3%; Score 281.8; DB 18; Length 909;

Best Local Similarity 93.1%; Pred. No. 3e-82;

Matches 225; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 GATATGTAATGACCCAGTCTCACAATTCATGTCACATCATGAGACAGATATCACC 60

Db 408 gatatcgcatgaccagctccctaaatcaatgctccacgctcagtgagacaggtcagc 467

QY 61 ATCAGCTGCAAGGCCAGTCAGAGATGTACTACTGCTGTAGCCTGTATCAACAAAACCA 120

Db 468 attctctgcaagcgacgacgagatgtagtactgctgtagcctgtatcaacaaaacca 527

QY 121 GGGCAATCTCCTAAACCTGCTGATTTACTGGGCATCCAGCGGACATCGAGTCCCTGAT 180

Db 528 gggcaatctccctaaactactgattcaatcgacatccaccggaacacgagctccctgat 587

QY 181 CGCTTCACAGCAGTGTGATCTGGACAGATTATCTCTACACATCAGCAGTGTGAGGCT 240

Db 588 cgttcacagcgacgtgagctcgggacagattatactctcacatcagcagtgctcaggt 647

QY 241 GAGACCTGCGACATTATTAATCTGTCAGCAACATTATATCACTCCGTGGACGTTGGTGGA 300

Db 648 gaagacctggaactcattactgctcgaacatlatagagtgcggtacacgttcggaagg 707

QY 301 GGGACCAAGCTGAGAT 317

Db 708 gggaccaagctcgaagat 724

RESULT 4

AAT78879

ID AAT78879 standard; cDNA; 909 BP.

XX

AC AAT78879;

XX

DT 09-OCT-1997 (first entry)

XX

DE Single chain antibody 741F8 coding sequence.

XX

KM Production; single-chain; binding protein; antibody; eukaryote; virus;

KM transcription activator; promoter; expression; adenovirus; E1A; PCR;

KM polymerase chain reaction; amplification; primer; herpes simplex virus;

KM thymidine kinase; vector; enhancer; translation; heterologous; ss.

XX

OS Synthetic.

XX

XX Key Location/Qualifiers

FT CDS 3..755

FT /*tag= a

FT /product= single chain antibody 741F8

XX

PN US5631158-A.

XX

PD 20-MAY-1997.

XX

PE 25-OCT-1993; 93US-0143498.

XX

PR 25-OCT-1993; 93US-0143498.

PR 05-JUN-1995; 95US-0461184.

XX

PA (CREA-) CREATIVE BIOMOLECULES INC.

PI Dorai H, Oppermann H;

PI WPI: 1997-288577/26.

DR P-PSDB; AAW22400.

XX

PT Production of cell line for producing single-chain binding protein -

PT using construct containing DNA encoding viral transcription

PT activator protein

XX

PS Example 2; Column 29-32; 24pp; English.

XX

CC The invention relates to methods of increasing production of a

CC single-chain binding protein, especially a single chain antibody,

CC by generating eukaryotic cell lines containing DNA encoding either

CC a viral transcription activator protein that acts on and stimulates

CC a viral promoter controlling the expression of DNA encoding the

PN	US5753204-A.	
XX		
XX	19-MAY-1998.	
XX		
PF	05-JUN-1995;	95US-0461838
XX		

RESULT	6
AAV22598	
ID	AAV22598 standard; cDNA: 909 BP.
XX	
AC	AAV22598;
XX	
DT	13-JUL-1998 (first entry)
XX	
EE	DNA encoding single chain binding site molecule (sfv) 741F8

Query Match	85.18;	Score 271.6;	DB 16;	Length 360;
Best Local Similarity	90.98;	Pred. No. 4.6e-79;		
Matches 289; Conservative	0;	Mismatches 29;	Indels 0;	Gaps 0

RESULT	ID
10	AAQ90424
	standard; DNA; 321 BP

DE DNA encoding immunoglobulin light chain of anti-idiotypic antibody
DE against human anticancer antibody.

KW Antibody; cancer; CDR; heavy chain; light chain; immunoglobulin
KW complementarity determining region; ss.

OS Mus sp.

PN JP07101999-A.

PD 18-APR-1995.

PF 06-OCT-1993; 93JJP-0272950.

PR 06-OCT-1993; 93JP-0272950.

PA (HAGI/) HAGIWARA Y.

DR WPI: 1995-182987/24

XX
E

PT Novel anti-idiotypic antibody against an human anticancer monoclonal antibody - and DNA sequences encoding the antibody, useful in pharmacology, medicine and biochemical fields.

PS Claim 22; Page 5; 28pp; Japanese.

Query Match	84.1%	Score 268.4	DB 16	Length 321
Best Local Similarity	90.3%	Pred. No. 5e-78		
Matches 287	Conservative 0	Mismatches 31	Indels 0	Gaps 0

[illegible]

RESULT	11
AAV39345	
ID	AAV39345 standard; cDNA; 394 BP.

AC AAV39345;

DT 21-SEP-1998 (first entry)

DE Mouse anti-HM1.24 antibody light chain variable region encoding cDNA.

KW Mouse; human; humanised; anti-HM1.24 antibody; myeloma; FR; CDR;

KW framework region; complementarity determining region; ss

Mus sp. OS

FH	key	Location/Qualifiers
1	204	

```
FT      /*tag= a
```

XX
XX

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

XX
X

XX
XX

PA (CHUS) CHUGAI SEIYAKU KK.

PI	Koishihara Y,	Kosaka M,	Ohtomo T,	Ono K,	Tsuchiya M,
PI	Yoshimura Y;				

XX Mus sp.
OS JP07101999-A.
PN 18-APR-1995.
PD 06-OCT-1993; 93JP-0272950.
XX 06-OCT-1993; 93JP-0272950.
PR 06-OCT-1993; 93JP-0272950.
XX
XX (HAGI/) HAGIWARA Y.
XX WPI; 1995-182987/24.
DR P-PSDB; AAR74958.
XX
PT Novel anti-idiotype antibody against an human anticancer monoclonal
PT antibody - and DNA sequences encoding the antibody, useful in
PT pharmacology, medicine and biochemical fields.
XX
PS Claim 21; Page 5; 28pp; Japanese.
XX
CC AAQ90420-Q90424 are DNA molecules encoding possible light chains of a
CC new anti-idiotype antibody against a human anticancer monoclonal
CC antibody. This antibody contains in its heavy chain 3 complementarily
CC determining regions CDRI (AAR74929-R74931), CDRI2 (AAR74932-R74933) and
CC CDRI3 (AAR74936-R74939), this is also true of the light chain which has
CC its own CDRI (AAR74944-R74946 and AAR85774), CDRI2 (AAR74947-R74949) and
CC CDRI3 (AAR74950-R74954) The antibody and DNA encoding it are useful in
CC pharmacological, medical and biochemical fields.
XX
XX Sequence 321 BP; 83 A; 84 C; 75 G; 79 T; 0 other;
SO
Query Match 81.9%; Score 261.2; DB 16; Length 321;
Best Local Similarity 89.5%; Pred. No. 1.1e-75;
Matches 281; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
OY 1 GATATCGTAATGACCCAGTCTCACAAATTCATGTCACATGACAGAGAGATATCAC 60
DB 1 gacatctgtatgaccagctccacaatctcatgcccacatcagtagagagagagctcagt 60
OY 61 ATCACCTGCAAGGCGCAGTGTGATGCTGCTGAGCCTGGTATCAACAAACCA 120
DB 61 atcacctgcaagcgccagctcagatgtagtactctgtagcctgtatcaacagaaccca 120
OY 121 GGGCAATCTCTTAACATTGATTTACTGGGATCCACCGGACACTGGAGTCCCTGAT 180
DB 121 gggcaatctctttaaactactgcttactcgcacatccctacggtacacgtgagctccctgat 180
OY 181 CGGTTACAGGAGTGTGATGCTGAGATTTACTGTCACATCCATCAGCAGAGTGCAGGCT 240
DB 181 cggttacagagagtgatgctgagatcttactgctcaccatcagcagcgagctcagcgt 240
OY 241 GAAGACCTGGCAGCTTTTACTGTCAGACCAATATATCACTCCGCTGGTGGTGA 300
DB 241 gaagacctggcagcttattactctgtcagcaaatattatgctcctccacagcttgcgagct 300
OY 301 GGGACCAAGCTGGA 314
DB 301 gggaccaagctgga 314
RESULT 14
ID ABA94232 standard; DNA; 303 BP.
XX ABA94232;
XX
XX 13-MAR-2002 (first entry)
XX
XX Mouse 04 antibody kappa light chain 2 variable region DNA.
XX

KW Neuromodulatory; central nervous system; CNS; SH19M22; LYM 22; AKR4;
KW ebvH19M MS119D10; ebv H19M CB2b6; CB21E12; CB21E7; MS119E5; virucide;
KW antiparkinsonian; neuroprotective; nootropic; vulnerary; mouse; ds.
XX
OS Mus sp.
XX
XX Key Location/Qualifiers
FH 1..303
FT CDS /*tag= a
XX
XX WO200185797-A1.
XX
XX 15-NOV-2001.
XX
XX 30-MAY-2000; 2000WO-US14902.
XX
XX 10-MAY-2000; 2000US-0568351.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION RES.
XX
XX Rodriguez M, Miller DJ, Pease LR;
XX WPI; 2002-066596/09.
XX P-PSDB; ABB07185.
XX
XX Novel neuromodulatory agent (a human Igm monoclonal antibody),
XX promoting neurite outgrowth, regeneration, remyelination and
XX neuroprotection in central nervous system, useful to treat
XX post-infectious encephalomyelitis -
XX
XX Disclosure; Fig 44; 219pp; English.
XX
XX The invention provides a neuromodulatory agent (1) capable of promoting
XX neurite outgrowth, regeneration, remyelination and neuroprotection in
XX central nervous system (CNS). (1) is capable of inducing remyelination,
XX promoting cellular proliferation of glial cells, and promoting Ca2+
XX signaling with oligodendrocytes. An humanised antibody to (1) can be
XX selected from antibody SH19M22 (LYM 22), ebvH19M MS119D10, ebv H19M
XX CB2b6, AKR4, CB21E12, CB21E7 or MS119E5. (1) is useful for stimulating
XX remyelination of CNS axons, stimulating proliferation of glial cells in
XX CNS axons, or treating demyelinating disease of CNS in a mammal in need
XX of such therapy. (1) is capable of binding to structures and cells within
XX CNS. (1) is preferably useful for treating a demyelinating disease of CNS
XX of a mouse infected with strain DA of Theiler's murine encephalomyelitis
XX (TMEV) or for treating a human being having multiple sclerosis, or a
XX human or domestic animal with a viral demyelinating disease, or a post-
XX neural disease or CNS. (1) is also useful for an in vitro method of
XX stimulating the proliferation of glial cells from mixed cell culture.
XX (1) is also useful for stimulating remyelination of CNS axons. The
XX antibodies are useful for preventing infection by a bacterium, virus or
XX like pathogen that causes demyelination or other neurodegenerative
XX condition in a subject. Methods where (1) is administered to a patient
XX are useful for treating multiple sclerosis, Parkinson's disease,
XX Alzheimer's disease, amyotrophic lateral sclerosis (ALS), a viral
XX demyelinating disease, CNS diseases, and other conditions in the CNS
XX where nerves are damaged as by trauma. The present sequence represents
XX the mouse 04 antibody kappa light chain 2 variable region DNA.
XX
XX Sequence 303 BP; 76 A; 80 C; 72 G; 75 T; 0 other;
SO
Query Match 81.4%; Score 259.8; DB 24; Length 303;
Best Local Similarity 91.1%; Pred. No. 3.2e-75;
Matches 276; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
OY 1 GATATCGTAATGACCCAGTCTCACAAATTCATGTCACATGAGTAGAGAGATATCAC 60
DB 1 gacatctgtatgaccagctccacaatctcatgcccacatcagtagagagagagctcagt 60
OY 61 ATCACCTGCAAGGCGCAGTGTGATGCTGCTGAGCCTGGTATCAACAAACCA 120
DB 61 atcacctgcaagcgccagctcagatgtagtactctgtgagcctgtatcaacagaaccca 120

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 05:38:45 ; Search time 2643.23 Seconds
(without alignments)
1628.889 Million cell updates/sec

Title: US-09-824-286-5

Perfect score: 319
Sequence: 1 GATATCGTATGACCCAGTC.....AGGACCAAGCTGAGATCT 319

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estln:*
4: em_estln:*
5: em_estor:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	281.8	88.3	974	10	BF578595 602093306
2	276.2	86.6	960	10	BF584737 602098485
3	251.8	78.9	985	10	BF581877 602098343
4	249.4	78.2	841	10	BF581734 602095514
5	236.6	74.2	887	10	BF584391 602096577
6	235.6	73.9	680	10	BE376043 601229480
7	232.8	73.0	594	10	BE309592 601094848
8	231.4	72.5	345	9	AA710291 vts3a04_r
9	231.2	72.5	673	12	BH065089 RPCI-24-3
10	230.6	72.3	950	10	BF781701 602104224
11	224.6	70.4	918	10	BF135931 601781261
12	221.6	69.5	594	12	BH104821 RPCI-24-2
13	220	69.0	881	10	BE913824 601669294
14	214	67.1	899	10	BG973020 602838676
15	210.4	66.0	883	10	BF580610 602093730
16	196	61.4	471	9	AI233978 EST230666
17	195.6	61.3	864	10	BI111610 602895973

18	193.4	60.6	871	10	BG962566
19	183.8	57.6	867	10	BG754732
20	183.6	57.6	891	10	BG540787
21	183.6	57.6	992	10	BP976253
22	182	57.1	826	10	BG684416
23	182	57.1	957	10	BG341934
24	182	57.1	959	10	BG341941
25	180.4	56.6	417	9	AM406228
26	180.4	56.6	422	9	AM407904
27	180.4	56.6	714	10	BG757897
28	180.4	56.6	750	10	BG755394
29	180.4	56.6	1130	10	BF662874
30	180.4	56.6	1198	10	BF663671
31	179	56.1	711	10	BG546866
32	179	56.1	867	10	BG539961
33	178.8	56.1	431	9	AM406886
34	178.8	56.1	471	9	AM406294
35	178.8	56.1	886	10	BG756818
36	177.8	55.7	858	10	BG756326
37	177.4	55.6	927	10	BG536845
38	177.2	55.5	426	9	AM405900
39	177.2	55.5	493	9	AM405753
40	176.4	55.3	712	10	BG538665
41	176.2	55.2	813	10	BG431143
42	175.8	55.1	431	10	BI305095
43	175.8	55.1	721	10	BG686441
44	175.8	55.1	727	10	BG684242
45	175.6	55.0	444	9	AV686883

ALIGNMENTS

RESULT 1
LOCUS BF578595
DEFINITION 602093306F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4207632 5', mRNA sequence.
ACCESSION BF578595
VERSION BF578595.1 GI:11652410
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murine; Mus.
REFERENCE 1 (bases 1 to 974)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: LMA9770 row: 1 column: 01
High quality sequence stop: 711.
Location/Qualifiers
1..974
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4207632"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: colon; Vector: PCMV-SPOrt; Site: 1; Noti; Site-2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT

281 a 262 c 232 g 199 t

Query Match	78.2%;	Score 249.4;	DB 10;	Length 841;
Best Local Similarity	89.7%;	Pred. No. 1.5e-67;		
Matches 279;	Conservative	0;	Mismatches 31;	Indels 1;
				Gaps 1;

OY	7	GTAATGACCCAGTCTCACAAATTCATTGTGCCAATCAGTAGAGAGAACAGTATCACTCAACC	66
Db	77	GTGATGACCCAGTCTCACAAATTCATTGTGCCAATCAGTAGAGAGAACAGGTTCAGCATCAACC	136
OY	67	TGC AAGGCCACTCAGGATGTGACTACTGCTGTAGCCTGGTATCAACAAAACCGAGCCA	126
Db	137	TGC AAGGCCACTCAGGATGTGGGTACTGCTGTAGCCTGGTATCAACAGAAGCA - GGCA	195
OY	127	TCTCTTAACCTCTTGATTTTACTGGGCATTCACCCGGCACACTGAGAGTCCCTTATGGCTTC	186
Db	196	TCTCTTAACACTGATTATTACTGGGCATTCACCCGGCACACTGGAGTCCCTTATGGCTTC	255
OY	187	ACAGGCAGTGGATCTGGGACAGATTATACTCTCACCATTACGACAGTGTGCAGAGCTAAGAC	246
Db	256	ACAGGCATGTGATCTGGGACAGATTATCTCACTCTCACCATTAGCAATGTGAGTCTGAAGAC	315
OY	247	CTGCGACTTATTTACTGTGCAGCAACATTATATCACTCCGTGGACGTTGGTGGAGGAGACC	306
Db	316	TTGGCAGAGTATTTCTGTGCAGCAATAATAGCAGCTATCCTTCACCGTTCCGTCTCTGGAGACC	375
OY	307	AAGCTGAGGAT	317
Db	376	AAGCTGAGGCT	386

FEATURES	RESULT	5
LOCUS	BF584391	
DEFINITION	BF584391	887 bp mRNA linear EST 12-DEC-2000
ACCESSION	6020965577E1	NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4216434 5',
VERSION	BF584391	mRNA sequence.
KEYWORDS	BF584391.1	GI:11658109
ORGANISM	EST.	
SOURCE	house mouse.	
COMMENT	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE	1 (bases 1 to 887)	
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .	
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)	
	Unpublished (1999)	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgabs-remail.nih.gov	
	Tissue Procurement: Jeffrey E. Green, M.D.	
	cDNA Library Preparation: Life Technologies, Inc.	
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)	
	DNA Sequencing by: Incyte Genomics, Inc.	
	Clone distribution: MGC clone distribution information can be	
	found through the I.M.A.G.E. Consortium/LNL at:	
	http://image.lnl.gov	
	plate: LLM9793 row: 1 column: 19	
	High quality sequence stop: 695.	
	Location/Qualifiers	

```

source
1. .887
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="4216434"
/clone_id="NCI_CGAP_Co24"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: colon; Vector: pCMV-Sport6; Site: 1; Notif:
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by life
Technologies. Note: this is a NCI_CGAP library."

BASE COUNT      248 a
ORIGIN           228 c      207 g      204 t

```

Query Match	73.9%;	Score 235.6;	DB 10;	Length 680;
Best Local Similarity	84.4%;	Pred. NO. 3e-63;		
Matches 265; Conservative	0;	Mismatches 49;	Indels 0;	Gaps 0;

[illegible]

RESULT	7
BE309592	
LOCUS	594 bp
BE309592	mrna
	1 insert
	EST 26-OCT-2000

DEFINITION	601090484EF1.NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3489635 5'...
ACCESSION	BE303592
VERSION	BE303592.1
KEYWORDS	GI:9168025
SOURCE	EST
ORGANISM	house mouse, Mus musculus
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

REFERENCE
1 (pages 1 to 594)
Mammalia, Euthetia, Rodentia, Sciurognathi, Muridae, Mus.
AUTHORS
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: ccarbs-remail.nih.gov

Tissue Procurement: Iohar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: L146530 row: 0 column: 12
High quality sequence: stop: 591.

```

FEATURES
Source
Location/Qualifiers
1. .594
/organism="Mus musculus"
/strain="C57/B6"

```

```

158 a      149 c      143 g      144 t
NIH"
/db_xref="taxon:10090"
/clone_1lb="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pcwv-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by life technologies. Investigators
providing samples: lothar Hennigausen/Robin Humphreys,
NIH"

```

BASE COUNT	158	a	149	c	143	g	144	t
ORIGIN	NIH ^a							

FEATURES	Location/Qualifiers
source	1. .673

FEATURES	source
Plate: L1AM9809	row: 0 column: 05
High quality sequence stop: 617.	
Location/Qualifiers	
1..950	
/organism="Mus musculus"	
/strain="FVB/N"	
/db_xref="taxon:10090"	
/clone="IMAGE:4222684"	
/clone_1fb="NCI CGAP Kid14"	
/lab_host="DH10B (T1 phage-resistant)"	
/note="Organ: kidney; Vector: pCMV-SPO0R6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library. 1"	
BASE COUNT	252 a 226 c 247 g 224 t 1 others
ORIGIN	
Query Match	72.38; Score 230.6; DB: 10; Length 950;
Best Local Similarity	83.08; Pred. No. 1.3e-61;
Matches 263; Conservative	0; Mismatches 54; Indels 0; Gaps 0

Query Match	72.3%	Score 230.6	DB 10	Length 950
Best Local Similarity	83.0%	Pred. No. 1.3e-61		
Matches 263	Conservative	0	Mismatches 54	Indels 0
				Gaps 0
Qy 1	GATATCGTAATGACCCAGTCTCACAAATTCATGTGCACATCAGTAGGAGACAGTATAC	60		
Db 80	GACATTTGGCTGACCCAGTCTCAAAAATTCATGTGCACATCAGTAGGAGACAGGCGAC	139		
Qy 61	ATCACCTGGCAAGGCCAGTCAGAGATGTGACTACTGCTGTAGCCCTGGTATCACAAAAACA	120		
Db 140	ATCACCTGCAAGGCCAGTCAGAGATTTGTGTAAGTCTGTAGCCTGGTATCACAGAAACTA	199		
Qy 121	GGGCAATCTCCTAAACTTCTGATTTACTGGGCATCCACCCGGCACACTGGAATCCCTGAT	180		
Db 200	GGGCAAGGCTCTTAAACCACTAATTTATGTGACTCAACCGGCACACTGGAATCCCTGAT	259		
Qy 181	CGCTTCACAGGCAGTGGATCTGGGACAATTTACTCCACCATGACAGCATGTGGAGGCT	240		
Db 260	CGCTTCACAGGCAGGAGATCTGGGACAATTTACTCTCACCATTAACATATATACAAATCT	319		
Qy 241	GAAGACCTGGGCACTTTATTTACTGTAGGACAATTTATATCATCCCGTGGACCTTGGTGA	300		
Db 320	GAAGACCTGGGCAAGATTTATTTCTGTGTGACAGATTTGGAATTTCTCTCACGTTCCGAGGG	379		
Qy 301	GGGACCAAGCTGGAGAT	317		
Db 380	GGGACCAAGTTGGAAT	396		

RESULT	11
LOCUS	BF135931
DEFINITION	BF135931 918 bp mRNA linear EST 24-OCT-2000
ACCESSION	M0178126.F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4009430 5'
VERSION	BF135931
KEYWORDS	BF135931.1 GI:10974971
SOURCE	EST.
ORGANISM	house mouse. Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 918) NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: LTA95245 Row: 1 Column: 15

High quality sequence stop: 680.

FEATURES	Location/Qualifiers
source	1. .918

```

/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone_image="IMAGE:4009430"
/clone_id="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site_1: NOTI;
Site_2: Salt; transgenic model WNT-1, expression driven by
MMV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dT. Library constructed by Life Technologies.
Investigator provided samples: Gilbert Smith, NIH"

```

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPII-24. For BAC library availability, please contact pierre de jong (pdejong@tigr.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 245 Row: J Column: 7
 Seq primer: T7
 Class: BAC ends

```

FEATURES
SOURCE
Location/Qualifiers
1. 594
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-245J7"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pTRABAC1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC Library produced by Peter de Jong. The
library was cloned in the pTRABAC1 cloning vector at the
BamHI sites using MboI partially digested male C57Bl/6J
DNA."

```

BASE COUNT	167 a	104 c	140 g	183 t
ORIGIN				
Query Match		69.5%	Score 221.6:	DB 12;
Best Local Similarity		86.3%	Pred No. 7.2e-59;	Length 594;
Matches 245; Conservative		0;	Mismatches 39;	Indels 0;
				Gaps 0

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
Plate: LLM9771 row: 1 column: 07

High quality sequence stop: 639.
Location/Qualifiers

FEATURES
SOURCE
1. .883
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4208022"
/clone_id="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."

BASE COUNT 233 a 219 c 229 g 202 t
ORIGIN

Query Match 66.0%; Score 210.4; DB 10; Length 883;
Best Local Similarity 83.2%; Pred. No. 2.8e-55;
Matches 263; Conservative 0; Mismatches 51; Indels 2; Gaps 2;

```
OY 3 TATGTAATGACCCAGTCTCACAATTCATGTCCACATCAGTAGAGACAGTATCACCAT 62
   ||| | ||||| ||| ||||| ||| ||||| ||||| ||||| ||||| |||||
Db 74 TATTTTGATGACCCAGCTCCCAATTCCTG-CTGTATCAGCAGAGAGGTTACCAT 132

OY 63 CACCTGCAAGGCCAGTGCAGTGTACTGCTGTAGCTGTGTATCAACAAAACCAAG 122
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133 AACCTGCAAGGCCAGTGCAGTGTACTGCTGTAGCTGTGTATCAACAAAACCAAG 191

OY 123 GCAATCTCTTAACCTTCTGATTTACTGGGATCCACCCGACACACTGGAGTCCCTGATCG 182
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 GCAGTCTCTTAACCTGATATATCTATCATCCCAATCGTATACTGGAGTCCCTGATCG 251

OY 183 CTTGACAGGAGTGGATGGAGACGATATATCTCACCATCAGCAGTGTGCAAGCTGA 242
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 CTTGACAGGAGTGGACATGGAGGATTTTCACTTCCACATCAGCAGTGTGCAAGCTGA 311

OY 243 AGACCTGGACCTTATTAATGCTCAGCAACATTATATCACTCCGTGAGCTTCGGTGGAGG 302
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 AGACCTGGACCTTATTAATGCTCAGCAAGATTATAGTCTCTCGAGCTTCGGTGGAGG 371

OY 303 GACCAAGCTGGAGATC 318
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 372 CACCAAGCTGGAAATC 387
```

Search completed: August 20, 2002, 08:55:02
Job time: 11777 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 09:25:29 ; Search time 1815.27 Seconds

(without alignments)
3873.427 Million cell updates/sec

Title: US-09-824-286-6

Perfect score: 336
Sequence: 1 CTCGACGAGCTCAGACCTGCG.....GCCAAGGAGCAGCGTCACC 336

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank:
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	302.4	90.0	759	6	AX256292	AX256292 Sequence
2	292.8	87.1	354	10	AF178588	AF178588 Mus muscu
3	283.8	84.5	354	10	MUSAB	L48671 Mus musculu
4	283.4	84.3	363	6	AX084361	AX084361 Sequence
5	283.4	84.3	363	6	AX286236	AX286236 Sequence
6	283	84.2	354	10	MUSIGHARI	M36228 Mouse Ig he
7	279.8	83.3	357	6	AX256252	AX256252 Sequence
8	277	82.4	360	10	MUSIGHAX	AX256252 Sequence
9	276.8	82.4	350	10	S76531	M36217 Mouse Ig he
10	275.4	82.0	345	10	MUSIGMO520	S76531 Ig VH-anti-
11	274.4	81.7	345	10	MM060456	M76414 Mouse IgM c
12	274.2	81.6	342	10	AF006578	U60456 Mus musculu
13	273.6	81.4	351	6	AR083821	AF006578 Mus muscu
14	273.6	81.4	351	10	MMASMSIH	AR083821 Sequence
15	273.6	81.4	420	6	AR169919	X75089 M.musculu
16	273.2	81.3	331	10	AF072800	AR169919 Sequence
17	273	81.2	333	10	MUSIGKCLJ	AF072800 Mus muscu
18	272.2	81.0	363	10	MMACHVR8	M15233 Mouse IgM c
19	271	80.7	318	10	MMHGVF	X90883 M.musculu
20	271	80.7	1431	6	AX055011	X58646 Mouse reatr
21	271	80.7	1431	6	AX055012	AX055011 Sequence
22	270.8	80.6	351	10	AF110502	AX055012 Sequence
23	270.4	80.5	337	10	MUSIGHBL	AF110502 Mus muscu
24	269.8	80.3	304	6	AR083817	K00721 Mouse Ig ac
25	269.6	80.2	337	10	MUSIGHBJ	AR083817 Sequence
26	269.6	80.2	340	10	MUSIGHBM	K00719 Mouse Ig ac
27	269.6	80.2	341	10	MUSIGHBK	K00722 Mouse Ig ac
28	269.6	80.2	346	10	MUSIGHBO	K00720 Mouse Ig ac
29	269.6	80.2	348	10	MUSIGHBN	K00724 Mouse Ig ac
30	269.6	80.2	376	10	MMAPHOXH	K00723 Mouse Ig ac
31	269.6	80.2	401	10	MM053526	X15471 Mouse partI
32	269.6	80.2	668	6	A25750	U53526 Mus musculu
33	269.6	80.2	669	6	AR051417	A25750 Variable re
34	269.6	80.2	708	6	A25744	AR051417 Sequence
35	269.6	80.2	708	6	AR051414	A25744 Variable re
36	269.4	80.2	321	10	MMHIGVT	AR051414 Sequence
37	269	80.1	387	10	AF276284	X58649 Mouse reatr
38	268.2	79.8	304	10	AF118978	AF276284 Mus muscu
39	268	79.8	339	10	S82436	AF118978 Mus muscu
40	268	79.8	342	10	MUSIGHBR	S82436 Ig VH-H2 Im
41	267.8	79.7	342	10	MUSIGHBP	K00727 Mouse Ig ac
42	267.8	79.7	355	10	MM055482	K00725 Mouse Ig ac
43	267.6	79.6	291	6	AR083822	U55482 Mus musculu
44	267.6	79.6	291	6	AR100553	AR083822 Sequence
45	267.6	79.6	291	10	MMVM13	AR100553 Sequence

ALIGNMENTS

RESULT 1
AX256292
LOCUS AX256292 759 bp DNA
DEFINITION Sequence 62 from Patent WO0171005.
ACCESSION AX256292
VERSION AX256292.1 GI:16075164
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 759)
AUTHORS
Kischer, R., Rietmuller, G., Luterbuesse, R., Borschert, K.,
Kischer, R., Mayer, M. and Hofmeister, R.
TITLE
Multifunctional polypeptides comprising a binding site to an
epitope of the nkx2d receptor complex
JOURNAL
Patent: WO 0171005-A 62 27-SEP-2001;
Kufner, Peter (DE)
FEATURES
1..759
location/Qualifiers
source
BASE COUNT 188 a 186 c 201 g 184 t
ORIGIN

[illegible]

Db	190	ATGTCCACACTGAGCATGACGAACAACAACCTCCAGACCAAGTTTCTTAAATGAAAC	249
Qy	241	AGCTGTCAAACTGATGACACAGCCATCTACTACTGTGCCAGAGAGGGTTTC-----T	291
Db	250	AGTCTGACAGACTGATGACACACACCATGATCTACTGTCCAGAGATCATCTACTATGATT	309
Qy	292	ACGGTAGATTCTCATGTGACTACTAGTGGGGCCCAAGGAGACCAGGTTCACC	336
Db	310	ACGGCTTATGCTATGTGACTACTAGTGGGGCCCAAGGAGACCAGGTTCACC	354
RESULT	5		
AX286236			
LOCUS	AX286236	363 bp	DNA
DEFINITION	Sequence 2 from Patent EP1076089.		Linear
ACCESSION	AX286236		PAT 20-NOV-2001
VERSION	AX286236.1	GI:17046098	
KEYWORDS			
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Deckmyn, H. and Cauwenberghs, N.		
TITLE	Cell lines, ligands and antibody fragments for use in pharmaceutical compositions for preventing and treating haemostasis disorders		
JOURNAL	Patent: EP 1076089-A 2 14-FEB-2001;		
FEATURES	K. U. LEUVEN RESEARCH & DEVELOPMENT (BE)		
source	1..363	Location/Qualifiers	
primer_bind	/organism="Mus sp."		
	/db_xref="taxon:10095"		
	1..24		
V.region	/note="Vhback primer"		
misc_feature	1..363		
	/note="complementary determining region one"		
misc_feature	147..195		
	/note="complementary determining region two"		
misc_feature	291..330		
	/note="complementary determining region three"		
primer_bind	330..360		
	/note="Vhfor primer"		
BASE COUNT	93 a 91 c 95 g 84 t		
ORIGIN			
Query Match	84.3%;	Score 283.4;	DB 6;
Best Local Similarity	91.3%;	Pred. No. 1.3e-74;	
Matches 315;	Conservative 0;	Mismatches 21;	Indels 9;
		Gaps 1;	
Qy	1	CTGCAGGAGTACGACACCTGGCGCTGGTGGGCGCTCCACAGAGCCCTGCATCATCTGCACCT	60
Db	10	CTGCAGGAGTCTGGGACCTGGCGCTGGTGGGCGCTCCACAGAGCCCTGCATCATCTGCACCT	69
Qy	61	GTCCTCGGATTTCATTAAACAGCATGATGGTGTACACTGGGTTCCGCACAGCTCCAGAAAG	120
Db	70	GTCCTCGGATTTCATTAAACAGCATGATGGTGTACACTGGGTTCCGCACAGCTCCAGAAAG	129
Qy	121	GCTCTGAGAGTGGCTGGGAGTCTTTGGCTGGTGGAGACACAAATTAATTCGGCTCTC	180
Db	130	GCTCTGAGAGTGGCTGGGAGTCTTTGGCTGGTGGAGACACAAATTAATTCGGCTCTC	189
Qy	181	ATGTCCACACTGAGCATGACGAACAACATTCACAGACCAATTTCTTAAATGAAAC	240
Db	190	ATGTCCACACTGAGCATGACGAACAACATTCACAGACCAATTTCTTAAATGAAAC	249
Qy	241	AGTGTCAAACTGATGACACAGCCATCTACTACTGTGCCAGAGAGGGTTTC-----T	291
Db	250	AGTGTCAAACTGATGACACAGCCATCTACTACTGTGCCAGAGATCATCTACTATGATT	309
Qy	292	ACGGTAGATTCTCATGTGACTACTAGTGGGGCCCAAGGAGACCAGGTTCACC	336

Db 310 ACGGCTATGCTAGTGGGCGCCAGGACGACGTCACC 354
RESULT 6
MUSIGHAET
LOCUS
DEFINITION Mouse Ig heavy-chain mRNA V region, partial cds from hybridoma H61-15VH.
ACCESSION M36228
VERSION M36228.1
KEYWORDS GI:194839
SOURCE Mouse (strain Balb/c), cDNA to mRNA, from hybridoma H61-15.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Kavalier,J., Caton,A.J., Staudt,L.M., Schwartz,D. and Gerhard,W.
TITLE A set of closely related antibodies dominates the primary antibody response to the antigenic site CB of the A/PR/8/34 Influenza virus hemagglutinin
JOURNAL J. Immunol. 145, 2312-2321 (1990)
MEDLINE 90375932
COMMENT Draft entry and computer-readable sequence for [J. Immunol. (1990) in press] kindly submitted by J.Kavalier, 06-JUL-1990.
FEATURES
source
1..354
/organism="Mus musculus"
/strain="BALB/c"
/sub_species="domesticus"
/db_xref="taxon:10090"
/map="chromosome 12"
/cell_line="H61-15"
/tissue_type="hybridoma"
1..354
/gene="Igh"
/c1..>354
/codon_start=1
/product="Immunoglobulin heavy chain V-region"
/protein_id="AAA38067.1"
/db_xref="GI:194839"
/translation="QVQLKESGPELVAPSSQSLTTCVSGEISLTSTCVHWYKPPCKGLHWLVWAGSTNYNSALMSRLSISKNSQVFLKMSLQTDITAMYYCARGHYGYAMDYWGQSTVYS"
BASE COUNT 90 a 89 c 93 g 82 t
ORIGIN
Query Match 84.2%; Score 283; DB 10; Length 354;
Best Local Similarity 91.7%; Pred. No. 1.7e-74;
Matches 311; Conservative 0; Mismatches 25; Indels 3; Gaps 1;
Db 1 CTGCAGAGTCAAGACCTGGCTGGTGGCCCTCAGAGCGTTCATCACTTGCACT 60
||| |||||
Db 10 CTGAAGAGTAGGACCTGGCTGGTGGCCCTCAGAGCGTTCATCACTTGCACT 69
||| |||||
Qy 61 GTCTGTGGGTTTTCATTAAACGAGTATGTTACACTGGGTTCCGACGCTCCAGAAAG 120
||| |||||
Db 70 GTCTGTGGGTTTTCATTAAACGAGTATGTTACACTGGGTTCCGACGCTCCAGAAAG 129
||| |||||
Qy 121 GGTCTGAGTGGCTGGGAGTCAATTTGGGCTGGTGGAGCACAAATTAATTGGGCTCTC 180
||| |||||
Db 130 GGCTGGAGTGGCTGGGAGTCAATTAATGAGTGGGCTGGTGGAGCACAAATTAATTGGGCTCTC 189
||| |||||
Qy 181 ATGTCCAGACTGAACATCAACAGACAAATTCACAGACCAATTTCTTAATAATGAAC 240
||| |||||
Db 190 ATGTCCAGACTGAACATCAACAGACAAATTCACAGACCAATTTCTTAATAATGAAC 249
||| |||||
Qy 241 AGTCTGCAAACTGATGACAGACCACTACTGTGCGCAGAGAGGTTTCTACGCG---TA 297
||| |||||
Db 250 AGTCTGCAAACTGATGACAGACCACTACTGTGCGCAGAGAGGTTTCTACGCGCTAC 309
||| |||||

Qy 298 GATTCTATGACTACTGGGCGCCAGGACCAAGGTCAC 336
||| |||||
Db 310 TATGCTATGACTACTGGGCGCCAGGACCAAGGTCAC 348
||| |||||
RESULT 7
AX256252
LOCUS
DEFINITION Sequence 22 from Patent WO0171005.
ACCESSION AX256252
VERSION AX256252.1
KEYWORDS GI:16075153
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequence.
REFERENCE
AUTHORS Kufer,P., Riethmuller,G., Luterbuese,R., Borschert,K., Kischel,R., Mayer,W. and Hofmeister,R.
TITLE Multifunctional polypeptides comprising a binding site to an epitope of the nkq2d receptor complex
JOURNAL Patent: WO 0171005-A 22 27-SEP-2001;
Kufer, Peter (DE)
FEATURES
source
1..357
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 85 a 87 c 104 g 81 t
ORIGIN
Query Match 83.3%; Score 279.8; DB 6; Length 357;
Best Local Similarity 91.2%; Pred. No. 1.6e-73;
Matches 309; Conservative 0; Mismatches 27; Indels 3; Gaps 1;
Db 1 CTGCAGAGTCAAGACCTGGCTGGTGGCCCTCAGAGCGTTCATCACTTGCACT 60
||| |||||
Db 10 CTGCAGAGTCAAGACCTGGCTGGTGGCCCTCAGAGCGTTCATCACTTGCACT 69
||| |||||
Qy 61 GTCTGTGGGTTTTCATTAAACGAGTATGTTACACTGGGTTCCGACGCTCCAGAAAG 120
||| |||||
Db 70 GTCTGTGGGTTTTCATTAAACGAGTATGTTACACTGGGTTCCGACGCTCCAGAAAG 129
||| |||||
Qy 121 GGTCTGAGTGGCTGGGAGTCAATTTGGGCTGGTGGAGCACAAATTAATTGGGCTCTC 180
||| |||||
Db 130 GGCTGGAGTGGCTGGGAGTCAATTAATGAGTGGGCTGGTGGAGCACAAATTAATTGGGCTCTC 189
||| |||||
Qy 181 ATGTCCAGACTGAACATCAACAGACAAATTCACAGACCAATTTCTTAATAATGAAC 240
||| |||||
Db 190 ATGTCCAGACTGAACATCAACAGACAAATTCACAGACCAATTTCTTAATAATGAAC 249
||| |||||
Qy 241 AGTCTGCAAACTGATGACAGACCACTACTGTGCGCAGAGAGGTT---CTACGGTA 297
||| |||||
Db 250 AGTCTGCAAACTGATGACAGACCACTACTGTGCGCAGAGAGGTTTCTACGCGTAC 309
||| |||||
Qy 298 GATTCTATGACTACTGGGCGCCAGGACCAAGGTCAC 336
||| |||||
Db 310 GCCTGTGTTGTTACTGGGCGCCAGGACCAAGGTCAC 348
||| |||||
RESULT 8
MUSIGHAD
LOCUS
DEFINITION Mouse Ig heavy-chain mRNA V region, partial cds. H220-22VH.
ACCESSION M36217
VERSION M36217.1
KEYWORDS GI:194815
SOURCE Mouse (strain Balb/c), cDNA to mRNA, from hybridoma H220-22.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Kavalier,J., Caton,A.J., Staudt,L.M., Schwartz,D. and Gerhard,W.

SOURCE immunoglobulin-kappa.
Mus musculus (strain BALB/c, sub-species domesticus) female ascites
CDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 345)
AUTHORS Deng, J., Chua, M.M., Andrews, G.C. and Karush, F.
TITLE Primary B-cell response to neuropeptide Y and bovine pancreatic
polypeptide
JOURNAL Mol. Immunol. 29 (7-8), 847-856 (1992)
MEDLINE 92342150
FEATURES
location/Qualifiers
1..345
/organism="Mus musculus"
/strain="BALB/c"
/sub-species="domesticus"
/db_xref="taxon:10090"
/map="chromosome 12"
/sex="female"
/cell_line="DSP2.2"
/cell_type="hybridoma"
/tissue-type="ascites"
1..345
/gene="Igm"
1..345
/partial
/gene="Igm"
/codon_start=1
/product="mu-Immunoglobulin"
/protein_id="AA51159.1"
/db_xref="GI:197944"
/translation="LESGPGLVAPSSLSITCTVSGSLSYGVHWVROPGKLEML
GLVWAGSTNNSALMSRLSKSDNSKSYFLKMNLSIQTDFTAMVYCAASPLITYYAM
DYMGGTSVTVSS"
4..278
/gene="Igm"
279..283
/gene="Igm"
284..291
/gene="Igm"
292..294
/gene="Igm"
295..345
/partial
/gene="Igm"
BASE COUNT 86 a 90 c 84 g 85 t
ORIGIN
Query Match 82.0%; Score 275.4; DB 10; Length 345;
Best Local Similarity 91.3%; Pred. No. 3.3e-72;
Matches 304; Conservative 0; Mismatches 26; Indels 3; Gaps 1;

QY 304 ATGACTACTGGGCGCCAGAGCAGCGTACC 336
|||||
DB 304 ATGACTACTGGGCGCCAGAGCAGCGTACC 336
RESULT 11
LOCUS MM060456 345 bp mRNA linear ROD 10-JUL-1996
DEFINITION Mus musculus Ig anti-DNA heavy chain VDJ (VH Q52) mRNA, partial
cds.
ACCESSION U60456
VERSION U60456.1 GI:1407739
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 345)
AUTHORS Wlochow, M.K., Alexander, A.L., Pippen, A.M.M., Lefkowitz, J.B.,
Pisetsky, D.S. and Gillespie, G.S.
TITLE Anti-DNA induced in Preautimmune NZB/W Mice by Bacterial DNA
Immunization
JOURNAL unpublished
REFERENCE 2 (bases 1 to 345)
AUTHORS Wlochow, M.K., Alexander, A.L., Pippen, A.M.M., Lefkowitz, J.B.,
Pisetsky, D.S. and Gillespie, G.S.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1996) Research Service, Durham VA Med Center, 508
Fulton Box 151G, Durham, NC 27705, USA
FEATURES
location/Qualifiers
1..345
/organism="Mus musculus"
/strain="(NZB X NZW)F1"
/db_xref="taxon:10090"
/chromosome="12"
/clone="B cell hybridoma F4-1"
1..345
/gene="VH Q52"
<1..>345
/gene="VH Q52"
/note="Description: VH gene; heavy chain variable region"
/codon_start=1
/product="Ig heavy chain"
/protein_id="AAB03591.1"
/db_xref="GI:1407740"
/translation="QVQLKESGPPVLPAPSSLSITCTVSGSLSYGVHWVROPGK
LEMLGLVWAGSTNNSALMSRLSKSDNSKSYFLKMNLSIQTDFTAMVYCAASPLITYYAM
DYMGGTSVTVSS"
BASE COUNT 89 a 86 c 91 g 79 t
ORIGIN
Query Match 81.7%; Score 274.4; DB 10; Length 345;
Best Local Similarity 91.1%; Pred. No. 6.7e-72;
Matches 306; Conservative 0; Mismatches 21; Indels 9; Gaps 1;


```

OY 241 AGTCGCAAACTGATGACACAGCCATCTACTACTGTGTGCCAGAGAGGGTTCTACGGTAGAT 300
    |||||||  |||||||  |||||||  |||||||  |||
Db 250 AGTCGCAAACTGATGACACAGCCATCTACTACTGTGTGCCAAACAGG-----GGTG 300
OY 301 TCTATGACTACTGGGGCCAAAGGACACGGTCACC 336
    |||||||  |||||||  |||  |||  |||
Db 301 GCTATGACTACTGGGGCTCAAGAACTCTACTCACC 336

RESULT 12
AF006578 342 bp mRNA linear ROD 12-JUL-1997
LOCUS Mus musculus anti-DNA autoantibody heavy chain variable region
DEFINITION mna, partial cds.
ACCESSION AF006578
VERSION AF006578.1 GI:2253323
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 342)
AUTHORS Weller,S., Conde,C., Knapp,A.M., Levallois,H., Giffillan,S.,
Pasquali,J.L. and Martin,T.
TITLE Autoantibodies in mice lacking terminal deoxynucleotidyl
transferase: evidence for a role of N-addition in the
polyreactivity and in the affinities of anti-DNA antibodies
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 342)
AUTHORS Weller,S., Conde,C., Knapp,A.M., Levallois,H., Giffillan,S.,
Pasquali,J.L. and Martin,T.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-1997) Laboratory of Immunopathology, Institute of
Immuo-Hematology, Central Hospital, 1 Place de l'hospital,
Strasbourg 67000, France

FEATURES
Source location/Qualifiers
1..342
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/chromosome="12"
/cell_line="spleen hybridoma cell line 64.3A11"
<1..>342
/note="anti-DNA autoantibody; V-D-J region"
/codon_start=1
/product="IgM heavy chain variable region"
/protein_id="AAB62902.1"
/db_xref="GI:2253324"
/translation="LDIEYGPGLVAPSGSLSTCTVSGFSLTTYGVHWKPPGKLEW
LGVIMAGSTNYNSALMSRLSISKDMSQVFLKMSIQSDDTAMYYCARGSLITYAM
DHGRLVSRPRG"

BASE COUNT 84 a 87 c 89 g 82 t

ORIGIN
Query Match 81.6%; Score 274.2; DB 10; Length 342;
Best Local Similarity 91.2%; Pred. No. 7.7e-72;
Matches 291; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```

```

Db 181 ATGTCCAGACTGAGCATCAGCAAAAGCACTCCAGAGCCAAAGTTTCTTAAAAATGAC 240
OY 241 AGTCGCAAACTGATGACACAGCCATCTACTACTGTGTGCCAGAGAGGGTTCTACGGTAGAT 300
    |||||||  |||||||  |||||||  |||||||  |||
Db 241 AGTCGCAATCTGATGACACAGCCATCTACTACTGTGTGCCAGAGAGGATTATATCCTAT 300
OY 301 TCTATGACTACTGGGGCC 319
    |||||||  |||  ||
Db 301 GCTATGACTACTGGGGCTCAAGAACTCTACTCACC 319

RESULT 13
AR083821 351 bp DNA linear PAT 01-SEP-2000
LOCUS AR083821
DEFINITION Sequence 36 from patent US 5977316.
ACCESSION AR083821
VERSION AR083821.1 GI:10010592
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE
1 (bases 1 to 351)
AUTHORS Chatterjee,M., Foon,K.A. and Chatterjee,S.K.
TITLE Monoclonal antibody 1A7 and related polypeptides
JOURNAL Patent: US 5977316-A 36 02-NOV-1999;
FEATURES location/Qualifiers
Source 1..351
/organism="unknown"

BASE COUNT 95 a 87 c 88 g 80 t 1 others
ORIGIN
Query Match 81.4%; Score 273.6; DB 6; Length 351;
Best Local Similarity 88.4%; Pred. No. 1.2e-71;
Matches 297; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```

```

OY 1 CTGCAGAGTCAGACCTGCGCTGGTGGCGCCCTCAGAGAGCGTGTCCATCATCTTGCACT 60
    |||||||  |||||||  |||||||  |||||||  |||
Db 10 CTGAAGAGATCAGACCTGCGCTGGTGGCGCCCTCAGAGAGCGTGTCCATCATCTTGCACT 69
OY 61 GTCTCTGGGTTTTCATTATACAGACTATGCTGTACACTGGTTCGCCAGCCTCCAGGAAG 120
    |||||||  |||||||  |||  |||  |||
Db 70 GTCTCTGGGTTTCTCATTTATCCAGATATAGTATGCTGTACACTGGTTCGCCAGCCTCCAGGAAG 129
OY 121 GGTCTGAGTGGCTGGGAGTCATTGTTGGGCTGGTGGAGCAACAATTATATTCGGCTTC 180
    |||||||  |||||||  |||  |||  |||
Db 130 GGTCTGAGTGGCTGGGAGTCATTGTTGGGCTGGTGGAGCAACAATTATATTCGGCTTC 189
OY 181 ATGTCCAGACTGATGACACAGCCATCTACTACTGTGTGCCAGAGAGGGTTCTACGGTAGAT 300
    |||||||  |||||||  |||  |||  |||
Db 190 AAATCCAGACTGATGACACAGCCATCTACTACTGTGTGCCAGAGAGGGTTCTTAAAAATGAC 249
OY 241 AGTCGCAAACTGATGACACAGCCATCTACTACTGTGTGCCAGAGAGGGTTCTACGGTAGAT 300
    |||||||  |||||||  |||  |||  |||
Db 250 AGTCGCAAACTGATGACACAGCCATCTACTACTGTGTGCCAGAGAGGGTTCTACGGTAGAT 309
OY 301 TCTATGACTACTGGGGCCAAAGGACCAAGCTCACC 336
    |||||||  |||||||  |||  |||  |||
Db 310 GCTATGACTACTGGGGCTCAAGAACTCTACTCACC 345

RESULT 14
MMSWS1H 351 bp mRNA linear ROD 23-FEB-1994
LOCUS MMSWS1H
DEFINITION M. musculus (A.SW) mRNA for ASWS1 antibody heavy chain variable
region.
ACCESSION X75099
VERSION X75099.1 GI:414151
KEYWORDS antibody; variable region; VH region.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE	1 (bases 1 to 351)
AUTHORS	Monestier,M.
TITLE	Direct Submission
JOURNAL	Submitted (09-SEP-1993) M. Monestier, Center for Molecular Medicine and Immunology, One Bruce Street, Newark, NJ 07103-2763, USA
REFERENCE	2 (bases 1 to 351)
AUTHORS	Monestier,M., Losman,M.J., Novick,K.E. and Aris,J.P.
TITLE	Molecular analysis of mercury-induced antinuclear antibodies in H-2S mice
JOURNAL	J. Immunol. 152 (2), 667-675 (1994)
MEDLINE	94110621
FEATURES	Location/Qualifiers
source	1..351
	/organism="Mus musculus"
	/strain="A.SW"
	/db_xref="taxon:10090"
	/chromosome="12"
	/cell_line="ASWS1 hybridoma"
	/cell_type="hybridoma"
V_region	1..351
	/gene="VH ASWS1"
	/product="ASWS1 Heavy chain variable region"
gene	1..351
	/gene="VH ASWS1"
	<1..>351
CDS	/gene="VH ASWS1"
	/codon_start=1
	/product="ASWS1 heavy chain variable region"
	/protein_id="CA52990.1"
	/db_xref="GI:414152"
	/translation="QVQLKESGPELVAPSGSITCTYVSGFSLRYSVHWYRPPGKGL LEWGMGTGGENTDYNSALKSRSLISKDNSKQVFLKNNISLTQTDITAMY YCARGGYDYV YAMDYMQGTSVTVS"
BASE COUNT	95 a 87 c 88 g 80 t 1 others
ORIGIN	
Query Match	81.4%; Score 273.6; DB 10; Length 351;
Best Local Similarity	88.4%; Pred. No. 1.2e-71;
Matches 297; Conservative 0; Mismatches 39; Indels 0; Gaps 0;	
OY	1 CTGCAGGATGAGGACCTGGCGCTGGTGGCGCCCTACAGAGCGTCGCATCTGGCACT 60
Db	10 CTGAAGGAGTCAAGGACCTGGCGCTGGTGGCACCTTACAGAGCGCTGCATCATGACACT 69
OY	61 GTCCTGGGTTTTCATTAAACAGACTAGTGGTGTACACTGGGTTCGCCAGCCTTCAGAAAG 120
Db	70 GTCCTGGGTTCTCATTTATTCAGATATAGTGTACACTGGGTTCGCCAGCCTTCAGAAAG 129
OY	121 GGTCTGAGAGTGGCTGGGAGTCAATTTGGCGCTGGTGGACACACAATTAATTAATTCGCTCTC 180
Db	130 GGTCTTGGAGTGGCTGGGAGTGAATGATATGGGGGTGGTGGAAACAGACTAATTAATTCAGCTCTC 189
OY	181 ATGTCCAGACTGAACATCAACAGACAGACAACTTCCAGAGCCCAATTTCTTAAATAATGAAC 240
Db	190 AAATCCAACTGAGAGCATCAGCAGACAACTCCAGACCAAGTTTCTTAAATAATGAAC 249
OY	241 AGTCGCAAACTGATGACACAGCCATCTACTACTGTGCGACAGAGGGTTCTAAGGTAGAT 300
Db	250 AGTCGCAAACTGATGACACAGCCATCTACTACTGTGCGACAGAGGATGTACTACGACTAT 309
OY	301 TCTATGACTACTGGGGCCAAAGGACCAAGCGGCACC 336
Db	310 GCTATGACTACTGGGGTCAAGAACTCAGTCAACC 345
RESULT 15	
ARI69919	420 bp DNA linear PAT 17-DEC-2001
LOCUS	ARI69919 Sequence 3 from patent US 6291208.
DEFINITION	ARI69919
ACCESSION	ARI69919.1 GI:17907878
VERSION	
KEYWORDS	

SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 420)
TITLE	Amend, N.N., Barber, B.H., Gates, G.C., Caterina, J.E. and Kiehn, M.H.
JOURNAL	Chimeric antibodies for delivery of antigens to selected cells of the immune system
FEATURES	Patent: US 6291208-A 3 18-SEP-2001; Location/Qualifiers 1..420
SOURCE	/organism="unknown"
BASE COUNT	95 a 111 c 108 g 106 t
ORIGIN	
Query Match	81.4% Score 273.6; DB 6; Length 420;
Best local Similarity	89.8% Pred. No. 1,2e-71;
Matches 307; Conservative	0; Mismatches 29; Indels 6; Gaps 1;

	Query Matched	Best Local Similarity	81.4% ; Score 273.6 ; DB 6 ; Length 420 ;
	Matches 307 ;	Conservative 0 ;	Mismatches 29 ; Indels 6 ; Gaps
QY	1	CTGCAGAGTACGAGACCTGGCCCTGGTGGGCCCTCACAGAGACCTGTCCATCACTTGCAC	60
Db	70	CTGAAGGAGTACAGGACCTGGCCCTGGTGGGCCCTCACAGAGACCTGTCCATCACTTGCAC	129
QY	61	GTCTCTGGGTTTCAATTAACAGCTATATGTGTACACTGGGTTGGCCACGCTCCAGGAAAG	120
Db	130	GTCTCTGGGTTTCAATTAACAGCTATATGTGTACACTGGGTTGGCCACGCTCCAGGAAAG	189
QY	121	GGTCGGAGTGGCGGGGGAGTCATTTGGGCTGGTGGAGCAAAATTAATTAATTCGGCTCTC	180
Db	190	GGTCGGAGTGGCGGGGGAGTCATTATGGGCTGGTGGAGCAATTAATTAATTCGGCTCTC	249
QY	181	ATGTCCAGACACATCAACACAGAGACAATTCACAGAGCCAAATTTCTTAAAAAATGAC	240
Db	250	ATGTCCAGACACATCAACAGAGACAATTCACAGAGCCAAATTTCTTAAAAAATGAC	309
QY	241	AGTCGCAAACTGATGACACAGCCATCTACTCTGTGCCAGAGAGGGTTCTACGGTATG--	258
Db	310	AGTCGCAAACTGATGACACAGCCATCTACTCTGTGCCAGAGAGGGTTATGGTACTTAC	369
QY	299	----ATTCTATGACTACTGGGGGACAGGAGCAAGGCTTCCAC	336
Db	370	CACATATGCTATGACTACTGGGGGACAGGAGCAAGGCTTCCAC	411

```
Search completed: August 20, 2002, 09:25:31
Job time: 13321 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 09:36:01 ; Search time 363.1 Seconds
(without alignments)
1588.772 Million cell updates/sec

Title: US-09-824-286-6

Perfect score: 336
Sequence: 1 CTCGAGGAGTCAGGACCTG.....GCCAAGGAGCCAGGTACAC 336

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_032802:*

- 1: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	336	100.0	336	19	AA97441
2	302.4	90.0	759	24	AA597140
3	279.8	83.3	357	24	AA597129
4	275.2	81.9	360	22	AA78183
5	274	81.5	714	18	AA786310
6	274	81.5	1173	18	AA786312
7	273.6	81.4	351	20	AA31382
8	273.6	81.4	420	18	AA77852
9	271	80.7	1431	22	AA91017

10	271	80.7	1431	22	AA91018	Chimeric 4H6 anti-260P9 hybridoma VH
11	270.2	80.4	357	21	AA38902	MUSIGHAEI Mouse Ig
12	269.8	80.3	304	20	AA31378	VH NO2/12.4-VK NO1
13	269.6	80.2	669	14	AA037462	VH NO2/12.4-VK NO1
14	269.6	80.2	708	14	AA037459	MUSIGHAD mouse Ig
15	267.6	79.6	291	20	AA31383	18-2-3/TRY202' sin
16	267.2	79.5	720	17	AA36463	Single chain bind1
17	267.2	79.5	720	17	AA13739	18-2-3-/TRY59. A
18	267.2	79.5	729	11	AA055715	Coding sequence of
19	267.2	79.5	729	14	AA051541	18-2-3/TRY59 singl
20	267.2	79.5	729	17	AA36464	Single chain bind1
21	267.2	79.5	870	13	AA13740	Ox VH-hinge-VL ins
22	266.4	79.3	870	13	AA036982	Coding sequence of
23	265.6	79.0	724	14	AA051540	18-2-3-/TRY202'
24	265.6	79.0	724	11	AA005714	Ganglioside GD2 sp
25	264.6	78.8	411	22	AA786854	H-chain variable r
26	264	78.6	353	13	AA027175	Nucleotide sequenc
27	263	78.3	411	18	AA790984	Murine anti-Lewis
28	260.8	77.6	796	21	AA50583	Mab 2B6 heavy cha1
29	260	77.4	334	17	AA734097	Murine consensus 1
30	259.8	77.3	359	21	AA27332	MUSIGHAEI mouse IgM
31	259.6	77.3	292	20	AA31385	Anti-DNA antibody
32	259.4	77.2	369	18	AA743807	Mouse anti-human I
33	259	77.1	333	19	AAV03489	MUSAB M. musculus
34	258	76.8	292	20	AA31380	ECOR1-HindIII inse
35	258	76.8	293	18	AA231389	PUR.4124 insert en
36	258	76.8	737	18	AA794606	Nucleotide sequenc
37	258	76.8	737	18	AA76113	Truncated scfv ant
38	258	76.8	3922	22	AA78409	Two linked VHXS d
39	258	76.8	3922	22	AA78409	Scfv sequences in
40	256.4	76.3	699	16	AA076280	Scfv anti-lysozyme
41	256.4	76.3	852	11	AA004467	Light chain variab
42	256.4	76.3	889	13	AA021096	Plasmid pSV1VHD1.3
43	256.4	76.3	895	16	AA076272	
44	256.4	76.3	915	11	AA004466	
45	256.4	76.3	923	16	AA704171	

ALIGNMENTS

RESULT 1	AA797441	AA797441 standard; cDNA; 336 BP.
ID	AA797441	
AC	AA797441	
XX		
DT	21-MAY-1998	(first entry)
XX		
DE	Monoclonal antibody CP.B8 heavy chain variable region cDNA.	
XX		
KW	Cytokine receptor; gamma common chain; gc chain; human;	
KW	blocking agent; monoclonal antibody; CP.B8; immunological disease;	
KW	myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;	
KW	insulin-dependent diabetes; inflammatory bowel disease;	
KW	sympathetic ophthalmia; uveitis; allergy; asthma; infection;	
KW	graft versus host disease; psoriasis; immunosuppressive; therapy;	
SS.		
XX		
OS	Mus musculus.	
XX		
PN	WO9743416-A1.	
XX		
PD	20-NOV-1997.	
XX		
PF	09-MAY-1997;	97WO-US07870.
XX		
PR	10-MAY-1996;	96US-0017466.
XX		
PA	(BIOU) BIOGEN INC.	
XX		
PI	Benjamin CD, Burkly LC, Hession C, Whitty A;	
XX		

OY 301 TCATGACTACTGGGGCCAAAGACCAAGGTCAC 336
|||||
Db 313 gctatgactactgtgggccaaggaccacgctcacc 348

RESULT 3

AAS97129
ID AAS97129 standard; cDNA: 357 BP.

AC AAS97129;

DT 26-FEB-2002 (first entry)

DE Anti-NG2D hybridoma 6H7E7 variable heavy chain DNA.

XX Homo sapiens.
OS
XX
XX MO200171005-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 26-MAR-2001; 2001WO-EP03414.
XX
XX PR 24-MAR-2000; 2000EP-0106467.
XX
XX PA (KUFE/) KUFER P.
XX
XX PI Kufer P, Riettmueller G, Lutterbuese R, Borschert K, Kischel R;
XX PI Mayer M, Hofmeister R;
XX
XX DR WPI: 2002-055119/07.
XX P-PSDB; AAU72837.

PT Multifunctional polypeptides comprising binding sites that specifically
PT recognise extracellular groups of the NG2D receptor complex and
PT domains which function as receptors or ligands, useful for treating
PT cancers and infectious diseases -

XX
XX PS Example 3; Fig 16; 114pp; English.

XX The invention relates to a multifunctional polypeptide comprising a
CC domain with a binding site that specifically recognises an extracellular
CC group of the NG2D receptor complex and a second domain which functions
CC as a receptor or ligand. The polypeptide and its associated
CC polynucleotide are used for the preparation of a pharmaceutical
CC composition for the treatment of cancer, infections and/or autoimmune
CC conditions. The cancer may be a tumour of the head and neck, stomach,
CC oesophagus, colon, liver, intrahepatic bile ducts, pancreas, lung,
CC larynx, breast, ovary, cervix, prostate, kidney, testis, thyroid,
CC bladder or brain, or a melanoma, myeloma, sarcoma, leukaemia or lymphoma.
CC The infectious diseases can be caused by viruses, bacteria, fungi,
CC protozoa or helminths. The autoimmune diseases include multiple
CC sclerosis, Grave's disease, ankylosing spondylitis, acute anterior
CC uveitis, Goodpasture's syndrome, myasthenia gravis, insulin-dependent
CC diabetes mellitus, rheumatoid arthritis, pemphigus vulgaris and
CC autoimmune hepatitis. The sequences represent the NG2D receptor DNA, DNA
CC encoding the polypeptides of the invention and PCR primers used to
XX amplify the DNA sequences.

XX Sequence 357 BP; 85 A; 87 C; 104 G; 81 T; 0 other;

Query Match

Best Local Similarity 83.3%; Score 279.8; DB 24; Length 357;
Matches 309; Conservative 0; Mismatches 27; Indels 3; Gaps 1;

OY 1 CTGACGAGTCAGACACTGGCTGGGCGCCCTTCACAGACCTGTCATCTTGCACT 60
|||||
Db 10 ctgcagagagtcagagaccctggcctgtggtgcacctacagagcctgtccatcactgcact 69
OY 61 GTCTGTGGCTTTTCATTAAACGACTATGCTGTACCTGGTTGGCCAGCTCCAGGAAG 120
|||||
Db 70 gtctgtgggttttcatttaaccagcctatgtgtacactgtatgcgcagcctccagaag 129
OY 121 GGTCTGAGTGGCTGGGAGTCATTTGGCTGTGTGAACACAAATTAATTCGGCTTC 180
|||||
Db 130 ggtctgagagtgctgggagtcataatggtctgtgtgaacacaaatataatcgtctc 189
OY 181 ATGTCCACAGCTGACATCAACAGACAAATTCACAGACCCAAATTTCTTAAAAATGAAC 240
|||||
Db 190 atgtccacagctgagagtcataatgagacaaacccaaagacaaatgtttcttaaaatgat 249
OY 241 AGTCTGCAACTGATGACACAGCCATCTACTGTGTGCGCAGAGAGGTTT---CTACGGTA 297
|||||
Db 250 agtctgcaactgtatgacacagccatgtactactgtgcagaggggtgtacgaagggtgcg 309
OY 298 GATTCATGACTACTGGGGCCCAAGGACACAGGTCAC 336
+ + + + +
Db 310 gctgtgtgtgtactgtggccaaggaccacgctcacc 348

RESULT 4

AAH78183
ID AAH78183 standard; DNA: 360 BP.

AC AAH78183;

DT 26-NOV-2001 (first entry)

DE Nucleotide sequence of heavy chain variable region of antibody 2C4.

XX Monoclonal antibody 2C4; sialoadhesin factor-2; SAF-2; allergic rhinitis;
XX allergy; asthma; anemia; eczema; lymphoma; systemic mastocytosis;
XX Leukemia; eosinophil; ds.

XX Mus sp.

XX MO200166126-A1.

XX PD 13-SEP-2001.

XX PF 05-MAR-2001; 2001WO-US07193.

XX PR 07-MAR-2000; 2000US-0187595.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PA (SMIK) SMITHKLINE BEECHAM PLC.

XX PA (UTJO) UNIV JOHNS HOPKINS.

XX PI Abrahamson JA, Bochner B, Erickson-Miller CL, Kikly KK;
XX PI Schleimer R;

XX WPI: 2001-570749/64.

XX P-PSDB; AAG63986.

XX Novel monoclonal antibody specific for human sialoadhesin factor-2 for
PT diagnosis, prevention, treatment of allergy, asthma, eczema or diseases
PT such as lymphoma, leukemia or systemic mastocytosis, in a mammal -
XX
XX PS Disclosure; Fig 1; 35pp; English.

XX The present sequence encodes the heavy chain variable region of murine
CC monoclonal antibody 2C4. This antibody binds to human sialoadhesin
CC factor-2 (SAF-2). The antibody is useful for treating or preventing
CC allergic rhinitis, allergies, asthma, anemia, eczema or diseases such
CC as lymphoma, leukemia or systemic mastocytosis in a mammal. It is also
CC useful for detecting the presence of a cell, especially eosinophil in
CC a sample, by detecting binding of the antibody to SAF-2. The antibody

CC can be coupled to toxins, antiproliferative drugs or radionuclides to
CC kill cells in areas of excessive SAF-2 expression.
XX
SQ Sequence 360 BP; 94 A; 92 C; 88 G; 86 T; 0 other;

Query Match 81.9%; Score 275.2; DB 22; Length 360;
Best Local Similarity 90.1%; Pred. No. 8,6e-75;
Matches 308; Conservative 0; Mismatches 28; Indels 6; Gaps 1;

OY 1 CTGACGAGCTCAGAGCCTGCTGGTGGCCCTCAGAGAGCCTGTCCATCTTGCTGACT 60
|| |||||
Db 10 ctaagagagctcagagcctgctgctggtgagccctcagagcctgctcactgctcat 69
OY 61 GTCCTGGGTTTTCATTAACACGCTATGCTACACTGGGTTGCCAGCTCCAGAGAA 120
|| |||||
Db 70 gctctggtgttctcattcaaccatctatgctcactggtctgcagcctccaggaag 129
OY 121 GGTCTGAGTGGCTGGAGTCAATTTGGCTGGTGAAGCAAAATTAATTCGGCTTC 180
|| |||||
Db 130 ggtctggagtgctggagtaataatggtctggtggaagacaaataataatcgtctc 189
OY 181 ATGTCAGACTGACATCAACAGAGCAATTCAGAGCCAAATTTCTTAAATAATGAC 240
|| |||||
Db 190 atgtccagactgagcactcagcaagaacacacccagagccttcttcttaaaataaac 249
OY 241 ACTTCGCAACAGTATGACACACCATCTACTGTGCGCAGAGAGGTTCTA-----CG 294
|| |||||
Db 250 agctgcgaacactgactgacacagccctgtactctgtgcagagacgtagtgcctcat 309
OY 295 GTAGATTCTATGACTACTGAGGGGCAAGGACCACCGTCCAC 336
|| |||||
Db 310 tactattctatgataactcgtgggtcaagaaacctcagtcacc 351

RESULT 5
AAT86310
ID AAT86310 standard; DNA; 714 BP.

AC AAT86310;

DT 06-APR-1998 (first entry)

DE Single chain anti-distalognanglioside GD2 antibody 3G6-scfv.

KM Antibody construct; distalognanglioside; GD2; single chain Fv fragment;

KM scfv; tumour; neuroblastoma; osteosarcoma; soft tissue sarcoma;

KM tissue imaging; target delivery; toxin; streptavidin;

KM pro-drug converting enzyme; GD2-targeted lymphocyte; ss.

OS Synthetic.

FN WO9734634-A1.

PD 25-SEP-1997.

PE 20-MAR-1997; 97WO-US04427.

PR 20-MAR-1996; 96US-0013703.

PA (SLOK) SLOAN KETTERING INST CANCER RES.

PI Cheung NV, Guo H, Larson SM, Rivlin K, Sadelain M;

DR WPI; 1997-479996/44.

PT Recombinant single chain anti-distalognanglioside GD2 antibody -

PT useful to detect tumour cells expressing GD2 and to target
PT therapeutic agents, e.g. toxins, to such cells
PS Disclosure: Page 11-12; 31pp; English.
XX The present sequence encodes a recombinant single chain peptide,
CC

CC 3G6-scfv. The peptide is an antibody construct comprising the variable
CC regions of the heavy and light chains of an antibody against
CC distalognanglioside (GD2) as a single chain Fv fragment (scfv). GD2 occurs
CC in many tumours types including neuroblastoma, osteosarcomas and other
CC soft tissue sarcomas, medulloblastomas, high grade astrocytomas,
CC melanomas and small cell lung cancer. The peptide can be detectably
CC labelled, preferably with 99m-TC, for tissue imaging of cells expressing
CC GD2. It can also be used to target delivery of a therapeutic or
CC pre-therapeutic agent, such as a toxin, streptavidin or a pro-drug
CC converting enzyme, to cells expressing GD2. The peptide may further
CC comprise CD8 to facilitate the formation of GD2-targeted lymphocytes.
CC T cells containing the peptide can also be used to target GD2-producing
CC tumour cells.

SQ Sequence 714 BP; 174 A; 168 C; 205 G; 167 T; 0 other;
Query Match 81.5%; Score 274; DB 18; Length 714;
Best Local Similarity 90.2%; Pred. No. 2,6e-74;
Matches 305; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

OY 2 TGCAGAGTCAAGACCTGGCTGGTGGCCCTCAGAGAGCCTGTCCATCTTGCACTG 61
|| |||||
Db 368 tgaagagatcagagcctgctggtggtggtccctcagagcctgctcactgctgctg 427
OY 62 TCTCTGGGTTTTCATTAACACGCTATGCTACACTGGGTTGCCAGCTCCAGAGAA 121
|| |||||
Db 428 tctctggtgttctcattcaaccatctatgctcactggtgtgcagcctccaggaagg 487
OY 122 GCTCTGAGTGGCTGGAGTCAATTTGGCTGGTGAAGCAAAATTAATTCGGCTTC 181
|| |||||
Db 488 gctctgagtggtctggagtaataatggtctggtggaagcacaataataatcgtccta 547
OY 182 TGTCCAGACTGAACATCAACAGAGCAATTCAGAGCCAAATTTCTTAAATAATGACA 241
|| |||||
Db 548 tgtccagactgagcactcagcaagaacactccaagagccttcttcttaaaataaac 607
OY 242 GCTCTGCAACGATGACACAGCCATCTACTGCTGCCA---GAGAGGTTCTACGGTNG 298
|| |||||
Db 608 gctctgcaaacactgactgacacagccatgactactgctgctggtgggttaactacgct 667
OY 299 ATTCTATGACTACTGAGGGGCAAGGACCACCGTCCAC 336
|| |||||
Db 668 atgcttggactactcgtgggtcaagaaacctcagtcacc 705

RESULT 6
AAT86312
ID AAT86312 standard; DNA; 1173 BP.

AC AAT86312;

DT 06-APR-1998 (first entry)

DE Single chain anti-distalognanglioside GD2 antibody 3G6-scfv-streptavidin.

KM Antibody construct; distalognanglioside; GD2; single chain Fv fragment;

KM scfv; tumour; neuroblastoma; osteosarcoma; soft tissue sarcoma;

KM tissue imaging; target delivery; toxin; streptavidin;

KM pro-drug converting enzyme; GD2-targeted lymphocyte; ss.

OS Synthetic.

FN WO9734634-A1.

PD 25-SEP-1997.

PE 20-MAR-1997; 97WO-US04427.

PR 20-MAR-1996; 96US-0013703.

PA (SLOK) SLOAN KETTERING INST CANCER RES.

PI Cheung NV, Guo H, Larson SM, Rivlin K, Sadelain M;

DR WPI; 1997-479996/44.

PT Recombinant single chain anti-distalognanglioside GD2 antibody -

PT useful to detect tumour cells expressing GD2 and to target
PT therapeutic agents, e.g. toxins, to such cells
PS Disclosure: Page 11-12; 31pp; English.
XX The present sequence encodes a recombinant single chain peptide,
CC

/note="encodes streptavidin"

FT XX
XX
PN W09734634-A1.
XX
PD 25-SEP-1997.
XX
PF 20-MAR-1997; 97MO-US04427.
XX
PR 20-MAR-1996; 96US-0013703.
XX
XX (SLOK) SLOAN KETTERING INST CANCER RES.
XX
XX Cheung NY, Guo H, Larson SM, Rivlin K, Sadelain M;
XX PI WPI; 1997-479996/44.
XX DR
XX PT Recombinant single chain anti-distalogsanglioside GD2 antibody -
XX PT useful to detect tumour cells expressing GD2 and to target
XX PT therapeutic agents, e.g. toxins, to such cells
XX
XX
XX Disclosure; Page 13; 31pp; English.
XX
XX The present sequence encodes a recombinant single chain peptide,
XX CC 366-scFv-streptavidin. The peptide is an antibody construct comprising
XX CC the variable regions of the heavy and light chains of an antibody against
XX CC distalogsanglioside (GD2) as a single chain Fv fragment (scFv). GD2 occurs
XX CC in many tumours types including neuroblastoma, osteosarcomas and other
XX CC soft tissue sarcomas, medulloblastomas, high grade astrocytomas,
XX CC melanomas and small cell lung cancer. The peptide can be detectably
XX CC labelled, preferably with 99m-Tc, for tissue imaging of cells expressing
XX CC GD2. It can also be used to target delivery of a therapeutic or
XX CC pre-therapeutic agent, such as a toxin, streptavidin (e.g. present
XX CC sequence) or a pro-drug converting enzyme, to cells expressing GD2. The
XX CC peptide may further comprise CD8 to facilitate the formation of
XX CC GD2-targeted lymphocytes. T cells containing the peptide can also be used
XX CC to target GD2-producing tumour cells.
XX
SQ Sequence 1173 BP; 270 A; 334 C; 340 G; 229 T; 0 other;

Query Match 81.5%; Score 274; DB 18; Length 1173;
Best Local Similarity 90.2%; Pred. No. 3.2e-74;
Matches 305; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

QY 2 TGCAGGAGTCAGACGCTGGCGCTGGCGCCCTCACAGAGCGCTGCCATCTGCACCTG 61
DB 368 tgaagaggtcaagagccctgctgctgctccacagagccctgctccatcctgcacgt 427
QY 62 TCTCTGGGTTTTCATTAAACGAGCTATGCTGTACACTGGGTTCCGACGCTCCAGGAAAG 121
DB 428 tctctgggtttcatttaaccagctatgtgtacactggttcgcacgctccacaggaag 487
QY 122 GTCTGGAGTGGCTGGGAGTCAATTGGGCTGCTGGAACACAAATTATATTCGGCTCTCA 181
DB 488 gtctggagtgctgggagatatactggtctggtgaagcacaaattataatctgcctcta 547
QY 182 TGTCCAGACTGACATCAACAGACAAATTCACAGAGCAAAATTTCTTAAATGAACA 241
DB 548 tgtccaagactgagatcatcagaagacaaccccaagagccaagtttcttaaaatgaaca 607
QY 242 GTCTGCAACTGATGACACAGCCATCTACTACTGTGCCA--GAGAGGGTTCTACGGTAG 298
DB 608 gtctgcaaaactgatacagacagccatgctactctgtgccacgcggggggttaactacgct 667
QY 299 ATTCTATGAGCTACTGGGCCAAGGACCAAGCGTCAAC 336
DB 668 atgcttgactactggtggttaagaaagaaactcaagtcacc 705

RESULT 7
AAZ31382
ID AAZ31382 standard; DNA: 351 BP.
XX

AC AAZ31382;
XX
DT 07-FEB-2000 (first entry)
XX
DE MMASWSIH M. musculus A.SW mRNA for ASWSI antibody heavy chain.
XX
KW Monoclonal antibody; Mab; 1A7; GD2; immune response; melanoma;
XX tumor-associated antigen; ss.
XX
OS Mus sp.
XX
PN US5977316-A.
XX
PD 02-NOV-1999.
XX
PF 16-JAN-1996; 96US-0591196.
XX
PR 17-JAN-1995; 95US-0372676.
XX
XX (KENT) UNIV KENTUCKY.
XX
XX Foon KA, Chatterjee SK, Chatterjee M;
XX
XX WPI; 1999-619711/53.
XX
XX Monoclonal antibody 1A7 which elicits an anti-GD2 immunological
XX PT response, useful for the development of products for the detection and
XX PT treatment of cancers -
XX
XX Disclosure; Fig 13B; 74pp; English.
XX
XX The invention provides a monoclonal antibody (Mab) designated 1A7, which
XX CC elicits an anti-GD2 (tumor-associated antigen) immunological response in
XX CC humans. Mab 1A7 has defined light and heavy chain variable region
XX CC sequences. The Mab 1A7 and polypeptides can be used for eliciting an
XX CC anti-GD2 immune response. The polypeptides can also be used for
XX CC detecting or purifying anti-GD2 antibody. The products can be used for
XX CC treating GD2-associated diseases, e.g. melanoma, neuroblastoma, glioma,
XX CC soft tissue carcinoma, and small cell carcinoma. They can be used for
XX CC palliating the disease or for reducing the risk of recurrence. Sequences
XX CC AAZ31373-90 represent sequences that closely matched Mab 1A7 heavy chain
XX CC variable region encoding sequence.
XX
SQ Sequence 351 BP; 95 A; 87 C; 88 G; 80 T; 1 other;

Query Match 81.4%; Score 273.6; DB 20; Length 351;
Best Local Similarity 88.4%; Pred. No. 2.6e-74;
Matches 297; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 CTGACGAGTCAGACACCTGGCGCTGGCGCCCTCACAGAGCGCTGCCATCTGCACCTG 60
DB 10 ctgaagaggtcaagagccctgctgctgctccacagagccctgctccatcctgcacgt 69
QY 61 GTCTCTGGGTTTTCATTAAACGAGCTATGCTGTACACTGGGTTCCGACGCTCCAGGAAAG 120
DB 70 gtctctggtttcattataaccagctatgtgtacactggttcgcacgctccacaggaag 129
QY 121 GGTCTGAGTGGCTGGGAGTCAATTGGGCTGCTGGAACACAAATTATATTCGGCTCTC 180
DB 130 ggtctgagtgctgggagatatactggtctggtgaagcacaaattataatcagctctc 189
QY 181 ATGTCCAGACTGACATCAACAGACAAATTCACAGAGCAAAATTTCTTAAATGAAC 240
DB 190 aaatccagactgagatcatcagaagacaaccccaagagccaagtttcttaaaatgaac 249
QY 241 AGTCTGCAACTGATGACACAGCCATCTACTACTGTGCCAAGAGGGTTCTACGGTAGAT 300
DB 250 agtctgcaaaactgatacagacagccatgctactctgtgccacgagatggttactaagactat 309
QY 301 TCTATGAGCTACTGGGCCAAGGACCAAGCGTCAAC 336
XXXXXXXXXXXXXXXXXXXX

DB	310	gctatggaactactgaggtcaaggaactcactgacc	345
RESULT	8		
AA77852			
ID	AA77852	standard; cDNA; 420 BP.	
XX			
AC	AA77852;		
XX			
DT	03-NOV-1997	(first entry)	
XX			
DE	Murine anti-human class II monoclonal antibody 44H104 HL chain cDNA.		
XX			
KW	Antibody; heavy chain; variable region; hybridoma cell line 44H104;		
KW	immune response; enhance; stimulate; vaccine; immunodiagnosis;		
KW	antigen delivery; ss.		
XX			
OS	Mus musculus.		
XX			
FT	Key	Location/Qualifiers	
FT	CDS	1..420	
FT		/*tag= a	
FT		/note= "Encodes 44H104 heavy chain variable region, including secretion signal; termination codon not given"	
XX			
PN	W09640941-A1.		
XX			
PD	19-DEC-1996.		
XX			
PP	07-JUN-1996;	96WO-CA00400.	
XX			
PR	07-JUN-1995;	95US-0483576.	
XX			
PA	(CONN-) CONNBUGHT LAB LTD.		
XX			
PI	Anand NN, Barber BH, Caterini JE, Cates GC, Klein MH;		
XX			
DR	WPI; 1997-077271/07.		
DR	P-PSDB; AAM22538.		
XX			
PT	Recombinant conjugate antibody mol., modified for delivering an		
PT	antigen - elicits enhanced immune response without the use of		
PT	adjuvant to generate antibodies which are useful in vaccines or		
PT	immunodiagnosis		
XX			
FS	Example 1; Fig 1B; 64pp; English.		
XX			
CC	Novel recombinant conjugate antibody molecules comprise a monoclonal		
CC	antibody specific for a surface structure of antigen presenting		
CC	cells (APC), genetically modified to contain at least one antigen		
CC	exclusively at one or more preselected sites. The conjugate is capable		
CC	of delivering the antigen to APC and eliciting an immune response to		
CC	the antigen. The new conjugates are useful as vaccines and are able		
CC	to elicit an enhanced immune response without the use of an adjuvant.		
CC	In a specific example, a conjugate was constructed using the murine		
CC	anti-human class II monoclonal antibody secreted by hybridoma		
CC	44H104. The peptide CLTB36 was chosen as antigen; it consists of		
CC	a tandemly linked T and B cell epitope derived from HIV MN strain.		
CC	The present sequence encodes the heavy chain variable region which		
CC	was PCR amplified from 44H104 and used in the preparation of a		
CC	conjugate with antigen CLTB36.		
XX			
XX			
SO	Sequence 420 BP; 95 A; 111 C; 108 G; 106 T; 0 other;		
Query Match	81.4%; Score 273.6; DB 18; Length 420;		
Best Local Similarity	89.8%; Pred. No. 2.8e-74;		
Matches 307; Conservative 0; Mismatches 29; Indels 6; Gaps 1;			
0Y	1	CTGCAGGAGTACAGACCTGGCTGGTGGCGCCCTACAGAGCGCTGTCATCATCTGGACT 60	
DB	70	ctgaaaggatcagagcccgccggtcggtgcgcctctacagagactgtccatccttgact 129	

```

Oy      61 GTCCTGGGTTTTCATTAAACAGACTAATGGTGTAACACTGGGTCCGCACGCTTCAGGAAG   120
Db      130 gctctgggttttcttaaccagcgcatacgtagtgacactggtgtccgcacgcttcagaaga   189
Oy      121 GGCTCGAGGTGGCTGGGAGTCAATTGGCTGGTGGAGCACAAATTATAATTGGGCTGC   180
Db      190 ggctcgaggatggctgggaatacatagggctggcgaaacaataattcaattcggctcc   249
Oy      181 ATGTCAGACTGAACATCACAACAGAGACAATTCCAAGAGACCAAATTTCTTAAAAATGAAC   240
Db      250 atgtcacgactgagcatcagcaagaacaactcaagagccaagtcttcttaaaaaatgagc   309
Oy      241 AGTCGCAACATCATGATGACACACACCACTCTACTCTGTGCCAGAGAGGGTTCTACGGTAG-- 298
Db      310 agtcgcgaacctgagcaacagcagcatactactctgtgccagagccatacgtgtactacgtc   369
Oy      299 ----ATTCTATGACCTACCTGGGGCCAGAGGACCAAGGTCAC   336
Db      370 cactatgtatgactactctgggtccaagaaacctcaagtcacc   411

RESULT          9
AAC91017
ID    AAC91017 standard; DNA; 1431 BP.
AC     AAC91017;
AD
AE
AF     15-MAR-2001 (first entry)
AG
AH
AI
AJ
AK
AL
AM
AN
AO
AP
AQ
AR
AS
AT
AU
AV
AW
AX
AY
AZ
BA
BB
BC
BD
BE
BF
BG
BH
BI
BJ
BK
BL
BM
BN
BO
BP
BQ
BR
BS
BT
BU
BV
BW
BX
BY
BZ
CA
CB
CC
CD
CE
CF
CG
CH
CI
CJ
CK
CL
CM
CN
CO
CP
CQ
CR
CS
CT
CU
CV
CW
CX
CY
CZ
DA
DB
DC
DD
DE
DF
DG
DH
DI
DJ
DK
DL
DM
DN
DO
DP
DQ
DR
DS
DT
DU
DV
DW
DX
DY
EZ
FA
FB
FC
FD
FE
FF
FG
FH
FI
FJ
FK
FL
FM
FN
FO
FP
FQ
FR
FS
FT
FU
FV
FW
FX
FY
FZ
GA
GB
GC
GD
GE
GF
GG
GH
GI
GJ
GK
GL
GM
GN
GO
GP
GQ
GR
GS
GT
GU
GV
GW
GX
GY
HA
HB
HC
HD
HE
HF
HG
HH
HI
HJ
HK
HL
HM
HN
HO
HP
HQ
HR
HS
HT
HU
HV
HW
HX
HY
HZ
IA
IB
IC
ID
IE
IF
IG
IH
II
IJ
IK
IL
IM
IN
IO
IP
IQ
IR
IS
IT
IU
IV
IW
IX
IY
IZ
JA
JB
JC
JD
JE
JF
JG
JH
JI
JJ
JK
JL
JM
JN
JO
JP
JQ
JR
JS
JT
JU
JV
JW
JX
JY
JZ
KA
KB
KC
KD
KE
KF
KG
KH
KI
KJ
KK
KL
KM
KN
KO
KP
KQ
KR
KS
KT
KU
KV
KW
KX
KY
KZ
LA
LB
LC
LD
LE
LF
LG
LH
LI
LJ
LK
LM
LN
LO
LP
LQ
LR
LS
LT
LU
LV
LW
LX
LY
LZ
MA
MB
MC
MD
ME
MF
MG
MH
MI
MJ
MK
ML
MM
MN
MO
MP
MQ
MR
MS
MT
MU
MV
MW
MX
MY
MZ
NA
NB
NC
ND
NE
NF
NG
NH
NI
NJ
NK
NL
NM
NO
NP
NQ
NR
NS
NT
NU
NV
NW
NX
NY
NZ
OA
OB
OC
OD
OE
OF
OG
OH
OI
OJ
OK
OL
OM
ON
OO
OP
OQ
OR
OS
OT
OU
OV
OW
OX
OY
PA
PB
PC
PD
PE
PF
PG
PH
PI
PJ
PK
PL
PM
PN
PO
PP
PQ
PR
PS
PT
PU
PV
PW
PX
PY
PZ
QA
QB
QC
QD
QE
QF
QG
QH
QI
QJ
QK
QL
QM
QN
QO
QP
QQ
QR
QS
QT
QU
QV
QW
QX
QY
RA
RB
RC
RD
RE
RF
RG
RH
RI
RJ
RK
RL
RM
RN
RO
RP
RQ
RR
RS
RT
RU
RV
RW
RX
RY
RZ
SA
SB
SC
SD
SE
SF
SG
SH
SI
SJ
SK
SL
SM
SN
SO

```


Db 67 ctgaagagtcacgacctggtcgtgctgcgccctcacagagcccttcatactgact 126
QY 61 GTCCTGGGTTTCATTAAACAGCTATGCTGACACAGGTTGGCCAGCCTCCAGGAAG 120
Db 127 gttctcgagtttcaataaccacctaagtgtacactggtctgcagcccaagaaag 186
QY 121 GGTCTGAGTGGCTGGGAGTCATTGGGCTGTGGAAGCACAATTTAAATTCGGCTCTC 180
Db 187 gttctcgagttgctggaataatagtggtctgtggaagacaaatataatctgctctc 246
QY 181 ATGTCCAGACTGAACATCAACAGACACATTTCCAAAGCCCAATTTTCTTAAAAATGAC 240
Db 247 atgtccagactgagcactcaagacaactccaagacaaagccaagttcttcttaaaatgaac 306
QY 241 AGTCTGCAAACTGATGACACAGCCATCTACTGTCGCCAGAGAGG 287
Db 307 agtctgcaaaactgatacagacgaatgactactcgtgcagaaagag 353

RESULT 10
AAC91018/c
ID AAC91018 standard; DNA: 1431 BP.

XX AAC91018;
XX 15-MAR-2001 (first entry)
XX DT
XX DE Chimeric 4H6 anti-DR4 antibody heavy chain complementary DNA.
XX KW Anti-Death receptor 4; DR4; antibody; apoptosis; cancer; arthritis;
XX KM autoimmune; ds.
XX OS Homo sapiens.
XX PN Synthetic.
XX PN WO200073349-A1.
XX PD 07-DEC-2000.
XX PF 25-MAY-2000; 2000WO-US14599.
XX PR 28-MAY-1999; 99US-0322875.
XX PA (GETH) GENENTECH INC.
XX PI Ashkenazi AJ, Chuntharapai A, Dodge KH, Kim KJ;
XX DR WPI; 2001-041145/05.
XX PT Novel anti-death receptor 4 antibodies useful for treating cancer and
XX PT immune related disorders such as rheumatoid arthritis, sjogren's
XX PT syndrome, Grave's disease and diabetes mellitus -
XX PS Disclosure; Fig 18; 126pp; English.
XX CC The present invention relates to an anti-death receptor 4 (DR4)
XX CC antibody. The antibodies of the invention are useful for inducing
XX CC apoptosis in mammalian cancer cells such as colon cancer cells and
XX CC for treating an immune-related disease in a mammal such as arthritis
XX CC and autoimmune disease.
XX SQ Sequence 1431 BP; 273 A; 367 C; 438 G; 350 T; 3 other;

Query Match 80.7%; Score 271; DB 22; Length 1431;
Best Local Similarity 96.5%; Pred. No. 2.9e-73;
Matches 277; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CTCGAGAGTCAGACCTGGCTGCTGTCGCGCCCTCACAGAGCCGTCATCACTTGACCT 60
Db 1365 CTCAGAGAGTCAGACCTGGCTGCTGTCGCGCCCTCACAGAGCCGTCATCACTTGACCT 1306
QY 61 GTCCTGGGTTTCATTAAACAGCTATGCTGACACTGCGGTGCGCCAGCCTCCAGGAAG 120

Db 1205 GTCCTGGGTTTCATTAAACAGCTATGCTGACACTGCGGTGCGCCAGCCTCCAGGAAG 1246
QY 121 GGTCTGAGTGGCTGGGAGTCATTTGGGCTGTGGAAGCACAATTTAAATTCGGCTCTC 180
Db 1245 GGTCTGAGTGGCTGGGAGTCATTTGGGCTGTGGAAGCACAATTTAAATTCGGCTCTC 1186
QY 181 ATGTCCAGACTGAACATCAACAGACACATTTCCAAAGCCCAATTTTCTTAAAAATGAC 240
Db 1185 ATGTCCAGACTGACATCAACAGACACATTTCCAAAGCCCAAGTTTCTTAAAAATGAC 1126
QY 241 AGTCTGCAAACTGATGACACAGCCATCTACTGTCGCCAGAGAGG 287
Db 1125 AGTCTGCAAACTGATGACACAGCCATCTACTGTCGCCAGAGAGG 1079

RESULT 11
AAA38902
ID AAA38902 standard; DNA: 357 BP.

XX AAA38902;
XX 29-AUG-2000 (first entry)
XX DE 260F9 hybridoma VH domain encoding DNA SEQ ID NO:13.
XX KW Antigen binding site; immunoglobulin; cancer antigen; immunological;
XX KM antibody; tumour; human; mucin; cancer; cytosolic; hybridoma;
XX KM specific binding assay; affinity purification; drug targeting;
XX KM toxin targeting; imaging; genetic; therapeutic; ss.
XX OS Homo sapiens.
XX PN US6054561-A.
XX PD 25-APR-2000.
XX PF 07-JUN-1995; 95US-0483749.
XX PR 21-MAR-1986; 86US-0842476.
XX PR 08-MAY-1988; 88US-0190778.
XX PR 08-FEB-1984; 84US-0577976.
XX PR 11-JAN-1985; 85US-0690750.
XX PR 11-AUG-1994; 94US-0288981.
XX PA (CHIR) CHIRON CORP.
XX PI
XX PI Ring DB;
XX DR WPI; 2000-338508/29.
XX DR P-PSDB; AAY90818.
XX PT Monoclonal antibody capable of binding to human breast cancer antigen
XX PT useful for affinity purification, drug or toxin targeting, imaging, and
XX PT treating cancer -
XX PS Disclosure; Fig 7; 57pp; English.
XX CC The present invention describes a monoclonal antibody (Mab) (I) that
XX CC binds to a human breast cancer antigen that is also bound by Mab 454C11
XX CC and 520C9 (produced by hybridoma ATCC HB8484 and HB8696, respectively). Also
XX CC described is a hybridoma that produces (I). (I) is useful in specific
XX CC binding assays, affinity purification, drug or toxin targeting, imaging,
XX CC and genetic or immunological therapeutics for various cancers. The
XX CC present sequence encodes a VH domain derived from a 260F9 hybridoma,
XX CC which is used in the exemplification of the present invention.
XX SQ Sequence 357 BP; 89 A; 89 C; 93 G; 86 T; 0 other;

Query Match 80.4%; Score 270.2; DB 21; Length 357;
Best Local Similarity 89.4%; Pred. No. 2.9e-73;
Matches 303; Conservative 0; Mismatches 33; Indels 3; Gaps 1;

[illegible]

Db	250	agctcgcaactgattgacacagcattgactactctgcagaggcatt	238
Db	250	agctcgcaactgattgacacagcattgactactctgcagaggcatt	238
RESULT	13		
AAQ37462			
ID	AAQ37462	standard; DNA; 669 BP.	
XX	AAQ37462;		
XX			
DT	19-JUN-1993	(first entry)	
XX			
DE	VH NQ2/12.4-VK NQ10/12.5	linked coding sequence #2.	
XX			
XX	Primer: human; immunoglobulin; Ig; variable region; VH; VL; Ck; JH;		
KM	lymphocyte; vector; soluble; antibody; phage; linker; back; VH3;		
KM	nested; in-cell PCR; cloning; polymorphic; TCR V; antiphenylloxazone		
KW	hybridoma; NQ2/12.4; NQ10/12.5; ss.		
XX			
OS	Synthetic.		
PH			
FT	Key	Location/Qualifiers	
FT	misc_RNA	1..339	
FT		/*tag= a	
FT		/label= VH_NQ2/12.4	
FT	primer_bind	1..23	
FT		/*tag= b	
FT		/label= NQ2/12.4_BACK	
FT	primer_bind	38..58	
FT		/*tag= c	
FT		/label= NQ2/12.4BKNS	
FT	primer_bind	complement (41..60)	
FT		/*tag= d	
FT		/label= NQ2HPRB	
FT	primer_bind	complement (307..352)	
FT		/*tag= e	
FT		/label= MOVHlnk4	
FT	misc_RNA	340..345	
FT		/*tag= f	
FT		/note= "linker peptide"	
FT	misc_RNA	346..669	
FT		/*tag= g	
FT		/label= VKappa_NQ2/12.4	
FT	primer_bind	327..370	
FT		/*tag= h	
FT		/label= MOVKappaInk4	
FT	primer_bind	complement (435..453)	
FT		/*tag= i	

FT	/label= NO2kappaPRB
FT	primer_bind
FT	complement (625..645)
FT	//tag= j
FT	/label= MOKappasFORNES
FT	primer_bind
FT	complement (647..669)
FT	//tag= k
FT	/label= MOKkappaSFOR2
XX	
PN	MO9303151-A.
XX	
XX	18-FEB-1993.
PD	
XX	
PF	10-AUG-1992;
XX	
PR	10-AUG-1991; 91GB-0017355-
XX	
PR	11-JUN-1992; 92GB-0012419.
PA	(MED1-) MEDICAL RES COUNCIL.
XX	
P1	Embleton MJ, Gorochov G, Jones PT, Winter GP;
DR	WPI: 1993-076508/09.
DR	P-PSDB: AAR32843.
XX	
PT	Treatment of cell populations, partic. hybridomas - to link
PT	together copies of 2 or more non-contiguous DNA sequences to
PT	facilitate analysis
XX	
PS	Disclosure: Fig 5; 72pp: English.
XX	
CC	The sequences given in AAQ37459-62 show the nucleotide sequences of
CC	mature heavy chain VH domains and the Vk light chain genes of the
CC	antiphenylloxazoline hydridomas NO2/12.4 and NQ10/12.5 which have been
CC	linked via a linker peptide by in-cell PCR. The cDNA was synthesised
CC	using forward primers annealing to the Ck gene and the JH segment,
CC	followed by assembly with linker primers, VH back primers based on the
CC	VH3 leader sequence and a forward CK primer nested in respect to the
CC	primer used for cDNA. The assembled product within the cells is then
CC	amplified with nested primers annealing to the 5' end of the VH gene
CC	and the 3' end of the Jk segment. In-cell PCR may be used to
CC	determine gene linkage analysis, particularly for the cloning of gene
CC	combinations that are polymorphic within a population of cells, such
CC	as the rearranged genes for Ig or TCR V regions.
SX	
SO	Sequence 669 BP; 165 A; 178 C; 174 G; 152 T; 0 other;
Query Match	80.2%; Score 269.6; DB 14; Length 669;
Best Local Similarity	96.8%; Pred. No. 5.8e-73;
Matches 275; Conservative	0; Mismatches 9; Indels 0; Gaps 0;
OY	1 CTGCAGGAGTCAGACCTGGCCTGGTGGCGGCCCTCACAGAGCGCTGTCCATCTTGCACT 60
OY	10 ctgaagaggtcagagccctgacctgttgcgccctcaacagagcttcataacttgacct 69
OY	61 GTCTGTGGGTTTCAATTAAACGACTATGCTGTACACTGGGTTCCGACGCTCCAGAAG 120
OY	70 gtccttgagtctcatcaaccagcatatggtgtacactggttcgcacgctccccaagaag 129
OY	121 GGTCTGGAGTGGCTGGAGTCTAATTGGGCTGGTGAAGAACACAAATTATAATTCCGGCTTC 180
OY	130 ggtctggagtggtcggtggagtaatataggctcgtgtgaaagacaacaattataatcggcttc 189
OY	181 ATGTCCACACTGACATCATCAACAGAGACAATTCCAAGACCATAATTTCTTAATAATGAAC 240
OY	190 atgtccaacactgagatcatcagcaaacacactccaagagccaagttctttaaaaatgaac 249
OY	241 AGTGTCAAACGTGATGACACACACCACTTACTACTCTGTGGCAGAGA 284
OY	250 agtctgcaaactgatgacacacagccatgtactactactgtgccagaga 293
Db	
RESULT	14

AAQ37459	standard; DNA: 708 BP.
AAQ37459	
19-JUN-1993	(first entry)
VH NQ2/12.4-Vk NQ10/12.5	linked coding sequence.
primer: human: immunoglobulin: Ig: variable region; VH; VL; Ck; JH;	
lymphocyte: vector: soluble; antibody: phage; linker; back; VH3;	
nested: in-cell PCR; cloning: polymorphic; TCR V; antiphenyloxazolone;	
hyridoma: NQ2/12.4; NQ10/12.5; ss.	
Synthetic.	
Key	Location/Qualifiers
misc_RNA	1..339
	/*tag= a
	/label= VH_NQ2/12.4
primer_bind	1..23
	/*tag= b
/label=	NQ2/12.4_back
primer_bind	38..58
	/*tag= c
/label=	NQ2/12.4BKNES
complement	(41..60)
/*tag= d	
/label=	NQ2HPRB
primer_bind	complement (322..371)
	/*tag= e
/label=	MOVhlnk3
misc_feature	340..384
	/*tag= f
/note=	"Linker Peptide"
	385..708
misc_RNA	
	/*tag= g
/label=	Vk_NQ10/12.5
	353..410
primer_bind	/*tag= h
/label=	MOVkappaInk3
primer_bind	complement (474..492)
	/*tag= i
/label=	NQ2KPRB
primer_bind	complement (664..684)
	/*tag= j
/label=	MOkappa5FORNES
primer_bind	complement (686..708)
	/*tag= k
/label=	MOUkappa5FOR2
MO9303151-A.	
18-FEB-1993.	
10-AUG-1992:	92MO-GB01483.
10-AUG-1991:	91GB-0017352.
11-JUN-1992:	92GB-0012419.
(MED1-) MEDICAL RES COUNCIL.	
Embleton MJ, Gorochov G, Jones PT, Winter GP;	
WPT: 1993-076508/09.	
P-PSDB: AAR32840.	
Treatment of cell populations, partic. hybridomas - to link	
together copies of 2 or more non-contiguous DNA sequences to	
facillitate analysis	
Disclosure: Fig 2; 72pp; English.	

CC The sequences given in AA037459-62 show the nucleotide sequences of
CC mature heavy chain VH domains and the VK light chain genes of the
CC antiphenylloxazone hydriodomas NQ2/12.4 and NQ10/12.5 which have been
CC linked via a linker peptide by In-cell PCR. The cDNA was synthesised
CC using forward primers annealing to the CK gene and the JH segment,
CC followed by assembly with linker primers, VH back primers based on the
CC VH3 leader sequence and a forward CK primer nested in respect to the
CC primer used for cDNA. The assembled product within the cells is then
CC amplified with nested primers annealing to the 5' end of the VH gene
CC and the 3' end of the JK segment. In-cell PCR may be used to
CC determine gene linkage analysis, particularly for the cloning of gene
CC combinations that are polymorphic within a population of cells, such
CC as the rearranged genes for Ig or TCR V regions.
XX
SQ Sequence 708 BP; 166 A; 184 C; 198 G; 160 T; 0 other;

Query Match 80.2%; Score 269.6; DB 14; Length 708;
Best Local Similarity 96.8%; Pred. No. 5.9e-73;
Matches 275; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CTGCAGAGTACAGACCTGGCTGGTGGCCCTCAGAGAGCTGTCCATCTGCACT 60
DB 10 ctgaagaggtcaagaccctggcctggtggccctcaccagagcctgtccatcactgcact 69
QY 61 GTCCTGGGTTTTCATTAAACAGCTATGATGACCTGGGTTCCGACCTCCAGGAAG 120
DB 70 gctctgggtttcatttaaccagctatggtgtacactggtgtcgcagcctccagaaag 129
QY 121 GGTCTGAGTGGCTGGGCTGATTTGGGCTGTGGAACACAAATTATTAATTTGGCTCTC 180
DB 130 ggtctggagtggtcgtggagtaataatggtcgtggagaaacacaataatcgtgcttc 189
QY 181 ATGTCACAGCTGAACATCAACAGACAAATTCACAAATTTCTTAAATGAAC 240
DB 190 atgtccagactgagatcagcaagaacaccccaagagccaggtttctttaaataatgaaac 249
QY 241 AGTCGCAAACTGATGACACAGCCATCTACTACTGTGCCAGCA 284
DB 250 agctcgcaactgatacagacagccatgtactactgtgccaagaa 293

RESULT 15
AAZ31383
ID AAZ31383 standard; DNA; 291 BP.
XX
AC AAZ31383;
XX
DN 07-FEB-2000 (first entry)
XX

DE MUSIGHAD mouse Ig heavy chain mRNA V region, partial CDS.

XX Monoclonal antibody; Mab; 1A7; GD2; immune response; melanoma;
KW neuroblastoma; glioma; soft tissue carcinoma; small cell carcinoma;
XX tumor-associated antigen; ss.

XX Mus sp.

XX US5977316-A.

XX 02-NOV-1999.

XX 16-JAN-1996; 96US-0591196.

XX 17-JAN-1995; 95US-0372676.

XX (KENT) UNIV KENTUCKY.

XX Foon KA, Chatterjee SK, Chatterjee M;

XX WPI; 1999-619711/53.

XX Monoclonal antibody 1A7 which elicits an anti-GD2 immunological

PT response, useful for the development of products for the detection and
PT treatment of cancers -
XX
XX Disclosure: Fig 13C; 74pp; English.

CC The invention provides a monoclonal antibody (Mab) designated 1A7, which
CC elicits an anti-GD2 (tumor-associated antigen) immunological response in
CC humans. Mab 1A7 has defined light and heavy chain variable region
CC sequences. The Mab 1A7 and polypeptides can be used for eliciting an
CC anti-GD2 immune response. The polypeptides can also be used for
CC detecting or purifying anti-GD2 antibody. The products can be used for
CC treating GD2-associated diseases, e.g. melanoma, neuroblastoma, glioma,
CC soft tissue carcinoma, and small cell carcinoma. They can be used for
CC palliating the disease or for reducing the risk of recurrence. Sequences
CC AAZ31373-90 represent sequences that closely matched Mab 1A7 heavy chain
CC variable region encoding sequence.
XX
SQ Sequence 291 BP; 76 A; 71 C; 77 G; 67 T; 0 other;

Query Match 79.6%; Score 267.6; DB 20; Length 291;
Best Local Similarity 96.8%; Pred. No. 1.7e-72;
Matches 273; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CTGCAGAGTACAGACCTGGCTGGTGGCCCTCAGAGAGCTGTCCATCTGCACT 60
DB 10 ctgaagaggtcaagaccctggcctggtggccctcaccagagcctgtccatcactgcact 69
QY 61 GTCCTGGGTTTTCATTAAACAGCTATGATGACCTGGGTTCCGACCTCCAGGAAG 120
DB 70 gctctgggtttcatttaaccagctatggtgtacactggtgtcgcagcctccagaaag 129
QY 121 GGTCTGAGTGGCTGGGCTGATTTGGGCTGTGGAACACAAATTATTAATTTGGCTCTC 180
DB 130 ggtctggagtggtcgtggagtaataatggtcgtggagaaacacaataatcgtgcttc 189
QY 181 ATGTCACAGCTGAACATCAACAGACAAATTCACAAATTTCTTAAATGAAC 240
DB 190 atgtccagactgagatcagcaagaacaccccaagagccaggtttctttaaataatgaaac 249
QY 241 AGTCGCAAACTGATGACACAGCCATCTACTACTGTGCCAGCA 282
DB 250 agctcgcaactgatacagacagccatgtactactgtgccaaga 291

Search completed: August 20, 2002, 09:36:03
Job time: 10478 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 08:55:02 ; Search time 2643.23 Seconds
(without alignments)
1715.695 Million cell updates/sec

Title: US-09-824-286-6
Perfect score: 336
Sequence: 1 CTCGAGGAGTCGACCTGTG.....GCCAAGGAGCACCAGCTCAC 336

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	270.2	80.4	840	10	BI690298
2	249	74.1	691	10	BF578188
3	238.2	70.9	671	10	BF182141
4	236.8	70.5	930	10	BF579344
5	226.4	67.4	685	12	BH275985
6	225.2	67.0	520	9	AW917371
7	219	65.2	637	10	BF531263
8	210.8	62.7	359	9	BB870665
9	198.6	59.1	425	10	BF578521
10	195	58.0	694	10	BF134274
11	174.4	51.9	507	10	BE845785
12	172.8	51.4	476	10	BE482206
13	172.8	51.4	496	10	BE485461
14	172.6	51.4	472	10	BE482203
15	171.4	51.0	414	10	BE478938
16	171.2	51.0	489	10	BE476735
17	170.8	50.8	511	10	BE485545

18	170.6	50.8	548	10	BG690980	339953	BA
19	170.2	50.7	558	10	BE845893	232632	BA
20	170	50.6	504	10	BE485128	172034	BA
21	169	50.3	523	10	BE483736	170020	BA
22	168.8	50.2	373	10	BE487845	176970	BA
23	168.6	50.2	413	10	BE481411	166849	BA
24	168.2	50.1	420	10	BF230480	252152	BA
25	167.8	49.9	497	10	BE486006	173340	BA
26	167.8	49.9	547	10	BG692467	342167	BA
27	167	49.7	500	10	BG692280	341915	BA
28	167	49.7	530	10	BG692277	341910	BA
29	167	49.7	549	10	BE845758	232868	BA
30	166.6	49.6	502	10	BE485907	173159	BA
31	166.4	49.5	505	10	BE589545	195790	BA
32	166.2	49.5	513	10	BG692944	342818	BA
33	166.2	49.5	602	10	BE588767	194388	BA
34	165.8	49.3	382	10	BE485343	172608	BA
35	165.4	49.2	501	10	BF230469	252126	BA
36	165.4	49.2	509	10	BE480721	165931	BA
37	164.8	49.0	513	10	BE483381	169533	BA
38	164.6	49.0	390	10	BE482316	168057	BA
39	164.6	49.0	412	10	BE482746	168642	BA
40	164.4	48.9	524	10	BE476249	158890	BA
41	164.2	48.9	376	10	BE486964	175386	BA
42	164	48.8	405	10	BE484472	171100	BA
43	164	48.8	511	10	BE478595	162919	BA
44	164	48.8	598	10	BE588685	194282	BA
45	163.6	48.7	381	10	BE488145	177466	BA

ALIGNMENTS

RESULT 1
BI690298 840 bp mRNA EST 18-SEP-2001
LOCUS 603309721F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5345741 5',
DEFINITION mRNA sequence.
ACCESSION BI690298 GI:15652927
VERSION BI690298.1 GI:15652927
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
REFERENCE 1 (bases 1 to 840)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: L1AM1878 row: 1 column: 06
High quality sequence stop: 828.
Location/Qualifiers
1. 840
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5345741"
/clone_lib="NCI CGAP Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies, Investigator
providing samples: Jeffrey Green, M.D., NIH*

Query Match	80.48;	Score 270.2;	DB 10;	Length 840;
Best Local Similarity	89.48;	Pred. No. 5.6e-72;		
Matches 303; Conservative	0;	Mismatches 33;	Indels 3;	Gaps 1;

Qy	1	CTGCAGAGATCAGACACCTGGCCCTGGTGGGGCCCTCAAGAGGCTGGTCCATCACTTGGACT	60
Db	85	CTGAAGAGATCAGAGACCTGGCCCTAGTGGCGCCCTACAGAGGCTGTGCATCACTATGGACT	144
Qy	61	GTCTCTGGGTTTCAATTAACACGATAGGTGTACACTGGTTCGGCAGCCTCCAGAAAG	120
Db	145	GTCTCTGGGTTTCTCAATTAAACGACTATGGTGTATGAGACTGGTTCGGCAGTCTCCAGAAAG	204
Qy	121	GATCTGAGATGGCTGGAGATCATTTGGGTGGGTGGAGACCAAAATATTAATTCGGTCTGC	180
Db	205	GSTCTGGAGTGGCTTGGAGTGTATATGGGGTGGTGGAGCACAAATATTAATTCAGCTTC	264
Qy	181	ATGTCACAGCTGATACATCAACAGAGACATTTCCAAAGAGCCAAATTTTCTTAAAAATGAAC	240
Db	265	AAATCAACACGTAGGATCAGCAAGACCAACTCCAAAGAGCCAAATTTTCTTAAAAATGAAC	324
Qy	241	AGTCTGCAAACTGATGACACAGCACACTTACTACTGTGCGAGAGAGAGGTT---CTAAGGTA	297
Db	325	AGTCTGCAAACTGATGACACACACCCTATCTACTCTGTCCAGTGGGGGGTCTACTAACGGAG	384
Qy	298	GATTTCTATGAGCTACTGGGGGCCCAAGGACACACGGTCAACC	336
Db	385	GATCTGATGAGCTACTGGGGGTCAAGGAACCTTCAATGTCACC	423

RESULT	2
LOCUS	Bf578188
DEFINITION	Bf578188 691 bp mRNA linear EST 12-DEC-2000 602094669J.FP1 NCI_CGAP_CO24 Mus musculus cDNA clone IMAGE:4209022 5'
ACCESSION	Bf578188
VERSION	Bf578188.1
KEYWORDS	GI:11651900
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 691)
AUTHORS	NIH-MGC http://mgc.nhl.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D.
COMMENT	

FEATURES	Location/qualifiers
source	1. .691

Technologies. Note: this is a NCI-CCGAP Library.

Query Match	74.1%;	Score 249;	DB 10;	Length 691;
Best Local Similarity	86.3%;	Pred. No. 1.7e-65;		
Matches 289;	Conservative 0;	Mismatches 40;	Indels 6;	Gaps 1.

QY	1	CTGCAGGAGATCAGGAGACCTGGCCCTGGTGGGGCCCTACAGAGCCTGTCATCTCACTTGCAC	60
Db	82	CTGAAGGAGATCAGGAGACCTGGCCCTGGTGGGGCCCTACAGAGCCTGTCATCTCACTTGCAC	141
QY	61	GTCCTGGGTTTTCATTAAACGACTATGTTGACACTGGGTTCCGCGACGCTCCAGGAAG	120
Db	142	GTCCTGGGTTTTCATTAAACGACTATGTTGACACTGGGTTCCGCGACGACCCAGGAAG	201
QY	121	GGTCTGGAGATGGCTGGGAGTCAATTTGGCGTGGTGAGACACAATATATATTCCTGGCTCTC	180
Db	202	GGTCTGGAGATGGCTTGGAGATATATGAGACTGGTGGAGGCACAAATATATATTCAGTCTTC	261
QY	181	ATGTCCAGACTAACAATCAACGAGACACATTTCCAGAGGCCAAATTTCTTAAATAATGAAC	240
Db	262	AAATCCAGACTGAGCATCAGCAAAACACTCCAAAGAGTCAAGTTTCTTAAATAATGAAC	321
QY	241	AGTTCGCAAACTGATGACACACGACACTTCTACTGCTGGCAGAGAGGTTCTTACAGGTAGAT	300
Db	322	AGTTCGCAAACTGATGACACACGACGATCTACTGTCGCCAAGAAATCGGCACTGGG-----	376
QY	301	TCTATGAGACTACTGGGGCCCAAGGACACAGGCTCAC	335
Db	377	-ACTTTGACTACTGGGGCCCAAGGACACAGGCTCTCAC	410

RESULT	3
BF182141	
LOCUS	671 bp mRNA linear EST 31-OCT-2000
DEFINITION	G01800468221 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4035429 5'
ACCESSION	BF182141
VERSION	BF182141.1 GI:11060283
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	1 (bases 1 to 671)
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)
	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.

FEATURES	Location/Qualifiers
source	1. .671

Library constructed by Life Technologies. Investigators
providing samples: Lochar Hennighausen/Robin Humphreys,
NIH

BASE COUNT 173 a 177 c 164 g 157 t
ORIGIN

Query Match 70.9%; Score 238.2; DB 10; Length 671;
Best Local Similarity 83.5%; Pred. No. 3,4e-62;
Matches 283; Conservative 0; Mismatches 53; Indels 3; Gaps 1;

QY 1 CTGCAGAGATCAGAGACCTGGCTGGGCGCCCTCAGAGACCTGTCCATCTGCACT 60
DB 67 CTGAAGAGATCAGAGACCTGGCTGGGCGCCCTCAGAGACCTGTCCATCTGCACT 126
QY 61 GTCCTGGGTTTCATTACCAAGCTATGGTACACTGGTTCGCCAGCTCCAGGAAG 120
DB 127 GTCTCGAGATTCATTACCAAGCTATGGTACACTGGTTCGCCAGCTCCAGGAAG 186
QY 121 GGTCTGAGTGGCTGGAGTCAATTTGGGCTGTGGAAGCACAATTAATTAATTCGCTTC 180
DB 187 GGTCTGAGTGGCTGGAGTCAATTTGGGCTGTGGAAGCACAATTAATTAATTCGCTTC 246
QY 181 ATGTCCAGACTGACATCAACAGAGCAATTCACAGAGCCAAATTTCTTAAAAATGAC 240
DB 247 ATGTCCAGACTGACATCAACAGAGCAATTCACAGAGCCAAATTTCTTAAAAATGAC 306
QY 241 AGTCTGAACTGATGACAGCAAGCTATCTACTGTCGAGCA---GAGGTTCTACGGTA 237
DB 307 AGTCTGAACTGATGACAGCAAGCTATCTACTGTCGAGCA---GAGGTTCTACGGTA 366
QY 298 GATTCTATGACTACTGGGGCCAGAGGACACGCTCACC 336
DB 367 TATGTATGACTACTGGGGCTCAGAGACCTCAGTACC 405

RESULT 4
LOCUS BF579344 930 bp mRNA linear EST 12-DEC-2000
DEFINITION 602095631F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4215751 5',
mRNA sequence.
ACCESSION BF579344
VERSION BF579344.1 GI:11653056
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 930)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM9791 row: n column: 08
High quality sequence stop: 633.
Location/Qualifiers
1..930
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4215751"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NciI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 240 a 237 c 249 g 204 t
ORIGIN

Query Match 70.5%; Score 236.8; DB 10; Length 930;
Best Local Similarity 81.5%; Pred. No. 1e-61;
Matches 274; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 CTGCAGAGATCAGAGACCTGGCTGGGCGCCCTCAGAGACCTGTCCATCTGCACT 60
DB 84 CTGAAGAGATCAGAGACCTGGCTGGGCGCCCTCAGAGACCTGTCCATCTGCACT 143
QY 61 GTCCTGGGTTTCATTACCAAGCTATGGTACACTGGTTCGCCAGCTCCAGGAAG 120
DB 144 GTCTCGAGATTCATTACCAAGCTATGGTACACTGGTTCGCCAGCTCCAGGAAG 203
QY 121 GGTCTGAGTGGCTGGAGTCAATTTGGGCTGTGGAAGCACAATTAATTAATTCGCTTC 180
DB 204 GGTCTGAGTGGCTGGAGTCAATTTGGGCTGTGGAAGCACAATTAATTAATTCGCTTC 263
QY 181 ATGTCCAGACTGACATCAACAGAGCAATTCACAGAGCCAAATTTCTTAAAAATGAC 240
DB 264 AAATCCAGACTGACATCAACAGAGCAATTCACAGAGCCAAATTTCTTAAAAATGAC 323
QY 241 AGTCTGAACTGATGACAGCAAGCTATCTACTGTCGAGCA---GAGGTTCTACGGTA 300
DB 324 AGTCTGAACTGATGACAGCAAGCTATCTACTGTCGAGCA---GAGGTTCTACGGTA 383
QY 301 TCTATGACTACTGGGGCCAGAGGACACGCTCACC 336
DB 384 ACTATGACTACTGGGGCTCAGAGACCTCAGTACC 419

RESULT 5
LOCUS BH275985 685 bp DNA linear GSS 30-NOV-2001
DEFINITION CH230-96J24, TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-96J24, DNA sequence.
ACCESSION BH275985
VERSION BH275985.1 GI:17188387
KEYWORDS GSS.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 685)
AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn,
A., Gebregregorys, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
Jong, P. and Fraser, C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 EcORI segment
JOURNAL Unpublished (1999)
COMMENT Other GSSs: CH230-96J24..TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@tigr.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_eing_information.htm). BAC end
page: http://www.tigr.org/tcdb/bac_ends/rat/bac_end_intro.html
Plate: 96 row: J column: 24
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1..685

FEATURES
source
Location/Qualifiers

FEATURES	SOURCE	location/qualifiers
		1..520
		/organism="Rattus norvegicus"
		/db_xref="taxon:10116"
		/clone="RGIED01"
		/clone_lib="Rat gene index, normalized rat, norvegicus"

FEATURES	Location/Qualifiers
source	1. 637
	/organism="Mus musculus"
	/strain="FVB/N"
	/db_xref="taxon:10090"
	/clone IMAGE:420516"
	/clone_11b="NCI_CGAP_Co24"

Yamanaka, A., Shibata, I., Hayatsu, N., Suganuma, I., Shibata, N., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

AUTHORS	Min-Moc Hsu: http://mgc.nci.nih.gov/
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@psi.barc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.960904.e. Vector identified by cross_match with the -mismatch 18 and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCAGC
Plate: 113 row: H column: 14
Seq primer: ATTTAGTGACACTATAG.
Location/Qualifiers

FEATURES
source 1.507
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."

BASE COUNT 109 a 153 c 135 g 110 t
ORIGIN

Query Match 51.9%; Score 174.4; DB 10; Length 507;
Best Local Similarity 76.4%; Pred. No. 1.1e-42;
Matches 214; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY 1 CTGCAGAGTCAGACCTGCTGTCGCGCCCTCAGAGAGCTGTCACACTTGCACCT 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 95 CTGCGGAGTCGCGGCCCGCCAGCTGCTGTAAGCCCTCAGACAGCCCTCCTCCACTGACG 154
OY 61 GTCTCTGGTTTCATTATTAACCACTATGCTACACTGGGTTGGCCAGCCTCCAGAAAG 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 155 GTCTCTGGATTCTCATTTGACACACTATGCTGTAATAAGGTCGCGCCAGCTCCAGGGAAG 214
OY 121 GGCTGTGAGTGGCTGGAGTCAATTTGGGCTGGTGAAGCAACAATTAATTTGGGCTCTC 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 215 GCTCTGGAGTGGCTTGGTGTATATATGCTGTCGAGAGCAGACTATATATCCGGCCCTG 274
OY 181 ATGTCCAGACTGAACATCAACAGACAAATTTCCAAAGCCAAATTTCTTAAAAATGAC 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 275 AATTCGCGACTCAGCATCACCAGCAAGACAGTTCCTCTGTGTCAGTGAGC 334
OY 241 AGTCTCAAACTGATGACACAGCCATCTACTGTGCCA 280
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 335 GGGCTGACACTGAGACAGCGCCACATCTACTGTGCCA 374

RESULT 12
BE482206 476 bp mRNA linear EST 28-AUG-2000
LOCUS BE482206
DEFINITION 167909 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE482206
VERSION BE482206.1 GI:9601739
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 476)
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
Wells,K.D.
TITLE Mapping of Expressed Sequence Tags from a normalized bovine mammary

JOURNAL Unpublished (2000)
COMMENT Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@psi.barc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.960904.e. Vector identified by cross_match with the -mismatch 18 and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCAGC
Plate: 22 row: O column: 5
Seq primer: ATTTAGTGACACTATAG.
Location/Qualifiers

FEATURES
source 1.476
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."

BASE COUNT 98 a 134 c 143 g 101 t
ORIGIN

Query Match 51.4%; Score 172.8; DB 10; Length 476;
Best Local Similarity 75.0%; Pred. No. 3.4e-42;
Matches 216; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

OY 1 CTGCAGAGTCAGACCTGCTGTCGCGCCCTCAGAGAGCTGTCACACTTGCACCT 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 22 CTGCGGAGTCGCGGCCCGCCAGCTGCTGTAAGCCCTCAGACAGCCCTCCTCCACTGACG 81
OY 61 GTCTCTGGTTTCATTATTAACCACTATGCTACACTGGGTTGGCCAGCCTCCAGAAAG 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 82 GTCTCTGGATTCTCATTTGACACACTATGCTGTAACAGTGGTCCGCGCCAGCTCCAGGGAAG 141
OY 121 GGCTGTGAGTGGCTGGAGTCAATTTGGGCTGGTGAAGCAACAATTAATTTGGGCTCTC 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 142 GCGCTGAGTGTCTTGGTGTATATATGCTGTCGAGAGCAGACTATATATCCAGCCCTG 201
OY 181 ATGTCCAGACTGAACATCAACAGACAAATTTCCAAAGCCAAATTTCTTAAAAATGAC 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 202 AATTCGCGCTCAGCATCACCAGCAAGACACTCCAAAGCCAAAGTCTCTGTGTCAGTGAGC 261
OY 241 AGTCTCAAACTGATGACACAGCCATCTACTGTGCCAAGAGGCT 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 262 AGCCTCAAACTGAGACAGCGCCACATCTACTGTGCCAAGAAAGT 309

RESULT 13
BE485461 496 bp mRNA linear EST 28-AUG-2000
LOCUS BE485461
DEFINITION 172501 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE485461
VERSION BE485461.1 GI:9604994
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 496)
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
Wells,K.D.
TITLE Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library

TS

USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200, Rm 2A, Beltsville, MD 20705, USA
Tel.: 301 504 8416
Tel.: 301 504 8416
Fax: 301 504 8414
Email: tade@lpsl.barc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980304.e. Vector identified by cross_match with the -mnscore 18
and -mismatch 12 options.

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_1b="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10a"
/notes="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."

```

Contact: Sonstegard TS

USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@lps1.barc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mismatches
and mismatch 12 options.

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_1lb="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV Sport6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."

```

Query Match	51.48;	Score 172.6;	DB 10;	Length 472;
Best Local Similarly	74.68;	Pred. No. 3.9e-42;		
Matches 217;	Conservative	0;	Mismatches 74;	Indels 0;
			Gaps	0;

Qy	1	CTGAGGAGCTAGGAGACCTGGCTGTGGGCTCTACAGAGACCTGTCCATCATCTATGAC	60
Db	160	CTGGGGAGTCGGGCCCCACCTGTGTAAAGCCCTCACAGACCCCTCTCCACCTGCAGC	219
Qy	61	GTCCTGTGGTTCATTATTAACGACGATGATGTATGCTACTGAGGTGGCCAGCCTCCAGAAAG	120
Db	220	GTCCTGTGATTCATCTATTAAAGGAGAAATATATGTAGGCTGGGGTCCGACGGCTCCAGGAAAG	279
Qy	121	GGTCGTGAGTGGCTGGGAGATCATTTTGGGCTGGTGGAAAGCACAAATTATATTTGGGCTCTC	180
Db	280	GGCGTGGAGTGGGCTGGGTGTCTATATATATGTGGTGGAAACACAGCTATTAACCCAGGCCCTG	339
Qy	181	ATGTCCACGATGGAATCATCAACAGAGACAATTCACAAAGACCCAAATTTTCTTAAAAATGAAC	240
Db	340	AAATCCCGGGCTCAGCATTCACCAAGGACAGGCTCCAAAGGCCAAGTCTCTATACACTGAGC	399
Qy	241	AGTCTGCAAACTGATGACACAGCCATCTACTACTGTCTCCAGAGAGGGTTCT	291
Db	400	AGCGTGAACACTGAGGACAGCGGCCACACTTACTCTGTCCAAAGGGGATTAAGTTGT	450

LOCUS	472 bp	mRNA	linear	EST
BE482203				28-AUG-2000
BE482203				
167903	5BOV	Bos taurus	CDNA 5', mRNA sequence.	

ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea
Bovidae; Bovinae; Bos.

AUTHORS	Sonstegard, T.S., Capuco, A.V., Van Tassel, C.P., Ashwell, M.S. and Wells, K.D.
TITLE	Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library
JOURNAL	Unpublished (2000)

Contact: Sonstegard TS

USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414

Email: tad@psi.barc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 18
and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCAGTCAGCAGC

Plate: 141 row: D column: 4

Seq primer: ATTAGTGACACTATAG.

Location/Qualifiers

1. 414

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_11b="BARC SBOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: PCMV SPORT6; Site.1: XbaI; Site.2: XhoI;

library made from pooled mRNA isolated from mammary

tissues at eight physiological, developmental, and disease

states."

BASE COUNT 88 a 125 c 116 g 85 t

ORIGIN

Query Match

Best Local Similarity 51.0%; Score 171.4; DB 10; Length 414;

Matches 214; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

OY 1 CTGCAGAGTCTGACCTGCTGGTGGGCGCCTCACAGAGCCGTCCATCAGTTCGACT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 CTGCGGAGTCTGCGGCCCGCCAGCTGTGAAGCCTTCACAGACCTCTCCCTGACG 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 61 GTCTGTGGGTTTCATTAAACAGCTATGTGTACACTGGGTGCCAGCCTCCAGAAAG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 GTCTGTGATTCATTGACAGCTATGTGTAACTGGGTGCCAGCCTCCAGGGAAG 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 121 GGTCTGAGTGGCTGGAGTCAATTGGGCTGGTGGGAAGCACAATTTAATTGGCTCTC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 GCGCTGAGTGGGTGGTGGCATCGGCACTAGTGAACACATCTATTACGACGCCCTG 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 181 ATGTCCAGACTGAACATCAACAGACCAATTTCCAAAGCCCAATTTCTTAAAAATGAC 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 AAATCCCGGCTCAGCATCACCAGACACTCCAAAGGCCAAGTCTCTGTGCAGTGAGC 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 241 AGTCTGCAAACTGATGACAGCAGCCATCTACTACTGTGCCAGAGAG 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 AGCGTGACACTGTAGACACAGCCACATCTACTGTGCAAGAGGG 398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: August 20, 2002, 08:55:05
Job time: 11780 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:13:26 ; Search time 94.22 seconds
(without alignments)
5.894 Million cell updates/sec

Title: US-09-824-286-3_COPY_28_32

Perfect score: 29

Sequence: 1 SYGVH 5

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	100.0	5	17	AA898945
2	29	100.0	83	22	AAU22812
3	29	100.0	83	22	AAU96118
4	29	100.0	107	13	AA828434
5	29	100.0	112	19	AAW31648
6	29	100.0	113	13	AA821268
7	29	100.0	115	18	AAW04595
8	29	100.0	116	11	AA807322
9	29	100.0	116	13	AA828287
10	29	100.0	116	17	AAW14450
11	29	100.0	116	17	AA898877

12	29	100.0	119	13	AA825728	Humanised VH regio
13	29	100.0	119	20	AA499218	Heavy chain variab
14	29	100.0	119	20	AA499808	K5B8 antibody heav
15	29	100.0	119	22	AA869650	Human Lay antibody
16	29	100.0	119	22	AA869675	Murine mIk-beta1 a
17	29	100.0	119	22	AA869676	Humanised mIk-beta
18	29	100.0	123	18	AAW07438	Anti-DNA antibody
19	29	100.0	127	11	AA806355	Peptide correspond
20	29	100.0	133	18	AAW10546	Wild type murine a
21	29	100.0	133	18	AAW10542	Humanised murine a
22	29	100.0	133	18	AAW10543	Humanised murine a
23	29	100.0	133	18	AAW10539	Humanised murine a
24	29	100.0	133	18	AAW10540	Humanised murine a
25	29	100.0	133	18	AAW10541	Humanised murine a
26	29	100.0	133	18	AAW10538	Humanised murine a
27	29	100.0	138	22	AA869656	Murine mIk-beta1 a
28	29	100.0	140	18	AAW22558	Murine anti-human
29	29	100.0	144	17	AA898944	Humanised anti-CD3
30	29	100.0	151	17	AA898941	Humanised anti-CD3
31	29	100.0	174	21	AA493723	The heavy chain of
32	29	100.0	188	22	AAE05352	Mouse secreted pro
33	29	100.0	219	15	AA856235	h66-118/h13-65/11-
34	29	100.0	222	14	AA822843	VH NQ2/12.4-VK NQ1
35	29	100.0	223	15	AA856236	166-111/112-65/h1-
36	29	100.0	235	14	AA832840	VH NQ2/12.4-VK NQ1
37	29	100.0	240	22	AAU33861	Staphylococcus aur
38	29	100.0	240	22	AAU36912	Staphylococcus aur
39	29	100.0	249	21	AA495527	Anti-CD38 antibody
40	29	100.0	260	21	AA495527	WOW-1 Fab heavy ch
41	29	100.0	264	16	AA895570	Intracellular bind
42	29	100.0	269	13	AA832569	Fusion protein enc
43	29	100.0	440	19	AAW42391	Thermococcus celer
44	29	100.0	450	21	AA44991	M79scfv-interleuk1
45	29	100.0	456	21	AA44992	M79scfv-interleuk1

ALIGNMENTS

RESULT 1

AA898945 ID AA898945 standard; Protein: 5 AA.

AA898945: AC

03-JAN-1997 (first entry)

CDR of humanised anti-CD38 monoclonal antibody heavy chain.

Monoclonal antibody: MAb; humanised; cancer; autoimmune disease;

multiple myeloma; lymphoma; rheumatoid arthritis; CD38;

complementary determining region; CDR; heavy chain; light chain;

ss.

Synthetic.

WO9616990-A1.

06-JUN-1996.

28-NOV-1995; 95WO-GB02777.

02-DEC-1994; 94GB-0024449.

(WELL) WELLCOME FOUND LTD.

Ellis JH, Lewis AP;

WPI; 1996-277724/28.

Humanised monoclonal antibodies with donor framework residues 29 and

78 - esp. against CD38, useful for treating cancer and auto-immune

diseases

XX Disclosure: Page 71, 92pp; English.
PS A monoclonal antibody (MAb) which has donor CDR's of foreign origin
XX and a recipient framework of human or primate origin, where the
CC original amino acid at position 29 or 78 of the heavy (H) chain of
CC the framework is replaced by an amino acid the same as or similar to
CC that in the corresponding position of the H chain of the Ab from
CC which the CDR's are derived, can be used for the treatment of cancer,
CC and autoimmune diseases, specifically multiple myeloma, lymphoma and
CC rheumatoid arthritis. The MAb binds to CD38. Replacing framework
CC residues 29 and 78 of the humanised Ab with the original donor
CC residues restores the antigen binding activity of the antibodies.
CC This complementary determining region of the heavy chain variable
CC region is designated CDRH1.
XX
XX Sequence 5 AA:
SQ

Query Match 100.0%; Score 29; DB 17; Length 5;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
Db 1 sygvh 5

RESULT 2
AAU22812
ID AAU22812 standard; Protein: 83 AA.
XX
XX AAU22812;
XX
XX 17-DEC-2001 (first entry)
XX
XX Human prostate cancer antigen, Seq ID No 331.
XX
XX
XX Human; prostate cancer antigen; cytostatic; uropathic; diagnostic;
KW reproductive system; chromosomal marker; forensic; urinary disorder;
KW chronic nephritis; blood-related disorder; thrombosis.
XX
XX Homo sapiens.
XX
XX MO200155316-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001MO-US01328.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218293.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227709.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244611.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249254.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barrash SC, Ruben SM;
XX
XX
DR MPI: 2001-451929/48.
DR N-PSDB; AAS40179.
XX
XX
PT Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the reproductive system including prostate cancer
PT and also for testing and detection e.g. diagnosis -
XX
XX
PS Claim 11; SEQ ID NO 331; 546bp; English.
XX
XX
CC The invention relates to novel isolated human prostate cancer antigen
CC polynucleotides (I) and polypeptides (II). (I) and (II) are useful for
CC preventing, treating or ameliorating a medical condition when
CC administered. (I), (II) and the antibody to (II) are useful for treating,
CC preventing and/or prognosing disorders related to the reproductive
CC system including prostate cancers; urinary disorders e.g. chronic
CC nephritis; and blood-related disorders e.g. thrombosis. (II) can be used
CC for testing and detection e.g. as a chromosomal marker and in forensics.
CC (I) and the anti-(II) antibody can be used in testing and detection in
CC immunoassays. AAU22702-AAU22913 represent the human prostate cancer
CC antigen amino acid sequences, and related amino acid sequences of the
CC invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SO Sequence 83 AA;

Query Match 100.0%; Score 29; DB 22; length 83;

Best Local Similarity 100.0%; Pred. No. 58;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYGVH 5
Db 16 sygvh 20

RESULT 3

AAAM96118
ID AAAM96118 standard; Protein; 83 AA.

AAAM96118;

21-NOV-2001 (first entry)

DE Human reproductive system related antigen SEQ ID NO: 4776.

KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.

OS Homo sapiens.
PN WO200155320-A2.

PD 02-AUG-2001.

PE 17-JAN-2001; 2001WO-US01339.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198122.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR XX
PR PA (HUMA-) HUMAN GENOME SCI INC.
PR XX
PR PI Rosen CA, Barash SC, Ruben SM;
DR WPI: 2001-465570/50.
DR N-PSDB: AAL02088.
PR XX
PR PT Isolated nucleic acid molecule encoding a reproductive system antigen
is used in preventing, treating or ameliorating a medical condition -
PR XX
PR BS Claim 11; SEQ ID NO 4776; 1297bp + Sequence Listing; English.
XX XX
CC The present invention provides the protein and coding sequences of a
number of human reproductive system related antigens. These can be used
in the prevention and treatment of reproductive system disorders,
including cancer. The present sequence is a protein of the invention.
CC XX
SQ Sequence 83 AA;

QY 1 SYGVH 5
 |||||
Db 16 sygvh 20

RESULT 4
AAR28434
ID AAR28434 standard; Protein: 107 AA.
XX
AC AAR28434;
XX
DT 04-APR-1993 (first entry)
XX
DE Sequence of monoclonal antibody NM-2 VH domain.
XX
KW Antibody: PCR primer; variable heavy chain;
anti-mucin monoclonal antibody.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Region 28..37
FT Region /label= CDR 1
FT Region 47..62
FT Region /label= CDR 2
FT Region 95..103
FT Region /label= CDR 3
XX
PN MO9218534-A.

XX 29-OCT-1992.
PD 92MO-GB00746.
XX
PF 23-APR-1992;
XX 91GB-0008652.
PR 23-APR-1991;
XX (ANTI-) ANTISOMA LTD.
PA
XX Courtenay-Luck NS;
PI
XX WPI, 1992-382045/46.
DR N-PSDB; AAQ30197.
XX
XX New peptide EPPT (Glu-Pro-Pro-Thr) - selectively binds mucin
PT expressed by epithelial tumours, used for guiding toxins or
PT labels to tumours expressing mucin
PS
XX Example; Table 4, page 30; 50pp; English.
XX
XX The antibody NM-2 is a murine monoclonal antibody class IgG.1,
CC Lambda light chain, which has specificity for the mucin mol.
CC The antibody reacts with about 95% of epithelial tumours and
CC cross-reacts with normal mucin. NM-2 VH and VL genes were
CC isolated using primers AAQ30193-Q30196. Both the heavy and the
CC light chain genes for NM-2 were sequenced and the CDR sequences
CC identified (see AAQ30197, AAQ30198).
XX
XX Sequence 107 AA;

Query Match 100.0%; Score 29; DB 13; Length 107;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SYGVH 5
| | | | |
DB 28 sygvh 32

RESULT 5
ID AAW31648 standard; Protein; 112 AA.
XX
XX AAW31648;
XX
XX 21-MAY-1998 (first entry)
XX
XX Monoclonal antibody CP.B8 heavy chain variable region.
XX
XX Cytokine receptor; gamma common chain; gc chain; human;
XX blocking agent; monoclonal antibody; CP.B8; immunological disease;
XX myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
XX insulin-dependent diabetes; inflammatory bowel disease;
XX sympathetic ophthalmia; uveitis; allergy; asthma; infection;
XX graft versus host disease; psoriasis; immunosuppressive; therapy;
XX complementarity determining region; CDR.
XX
XX Mus musculus.

XX Key Location/Qualifiers
XX Region 28..32
FT /note= "CDR1"
FT 47..61
FT /note= "CDR2"
FT 95..104
FT /note= "CDR3"
XX
XX MO9743416-A1.
XX
XX 20-NOV-1997.
XX
XX 09-MAY-1997; 97MO-US07870.

XX 10-MAY-1996; 96US-0017466.
PR (BIOJ) BIOGEN INC.
XX
XX Benjamin CD, Burkiy LC, Hession C, Whitty A;
PI
XX WPI; 1998-008885/01.
DR N-PSDB; AAT97441.
XX
XX Blocking agents of the gamma common chain of cytokine receptors -
PT particularly monoclonal antibodies, used to induce T cell anergy for
PT treatment of immunological diseases
XX
XX Claim 23; Page 81-82; 111pp; English.
XX
XX This polypeptide comprises the heavy chain variable region (VH) of
CC monoclonal antibody (Mab) C9.B8, which is produced by a hybridoma
CC deposited as ATCC 12107, and which is specific for the gamma
CC constant (gc) chain (see AAW31646) of human cytokine receptors. The
CC invention provides compositions and methods for inhibiting cytokine
CC signalling using gc chain blocking agents for the treatment of
CC immunological diseases such as myasthenia gravis, rheumatoid
CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,
CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,
CC allergy, asthma, parasitic infection, graft vs. host disease or
CC psoriasis. Preferred gc blocking agents include Mab CP.B8, its Fab
CC fragment and an antibody having a light chain variable region
CC CDR selected from those of CP.B8 VH or a heavy chain variable
CC region CDR selected from those of CP.B8 VL (see AAW31647).
XX
XX Sequence 112 AA;

Query Match 100.0%; Score 29; DB 19; Length 112;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SYGVH 5
| | | | |
DB 28 sygvh 32

RESULT 6
ID AAR21268 standard; Protein; 113 AA.
XX
XX AAR21268;
XX
XX 21-MAY-1992 (first entry)
XX
XX Murine VH group 2 chain E specific for phox.
XX
XX Fd; bacteriophage; gene III; filamentous; phagemid; capsid; coat;
XX plus; gpI; binding; adsorption; gene VIII; diverse repertoire;
XX specific binding pairs; replicable genetic display package.
XX
XX Synthetic.

XX Key Location/Qualifiers
XX Binding-site 31..35
FT /label= CDR1
FT 50..65
FT /label= CDR2
FT 98..102
FT /label= CDR3
FT /note= "D/N-X-G-X-X motif "
XX
XX MO9201047-A.
XX
XX 23-JAN-1992.
XX
XX 10-JUL-1991; 91MO-GB01134.

xx	15-MAY-1991;	91GB-0010549.	
pr	10-JUL-1990;	90GB-0015198.	
pr	19-OCT-1990;	90GB-0022845.	
pr	12-NOV-1990;	90GB-0024503.	
pr	06-MAR-1991;	91GB-0004744.	
xx	(CAMP-) CAMBRIDGE ANTIBODY.		
pa	(MEDI-) MED RES COUNCIL.		
xx			
pi	McCaferly J, Pope AR, Johnson KS, Hoogenboom HRJ, Griffiths AD;		
pi	Jackson RH, Holliger KP, Marks JD, Clackson TP, Chiswell DU;		
pi	Winter GP, Bonnett TP;		
xx			
dr	WPI; 1992-056862/07.		
xx			
pt	Producing members of specific binding pairs - by expression in		
pt	recombinant host cells with a secreting replicable genetic		
xx	display package.		
xx			
ps	Example 21; Fig 24; 209pp; English.		
xx			
cc	The VH sequence is one of eight (AAR21264-71) found to be expressed		
cc	from a single chain Fv library from an immunised mouse. The		
cc	library produces a diverse repertoire of antibody fragments specific		
cc	for 2-phenyl-5-oxazolone (phox). It was prep'd. using cDNA generated		
cc	from mRNA from mice immunised with phox coupled to chicked serum		
cc	albumin. The VH and VL kappa sequences were separately amplified by		
cc	PCR (see AAO23474-84) and ligated into fdCAT2 (see AAO23463) for		
cc	expression on the phage surface as fusions with gene III. The		
cc	resulting library of clones was diverse. Twenty three hapten binding		
cc	clones were sequenced revealing the eight different VH genes (A-H)		
cc	in a variety of pairings with seven different VK genes (a-g) (see		
cc	AAR21264-92). Nearly all the VH genes belonged to gp 1, with only		
cc	one, "E", being of gp 2 (Vhox1). Of the twenty three clones		
cc	sequenced, only one was of type "E". Most of the clones were VK-d		
cc	combinations. The Kd of VH-B/VK-d for phox-GABA was 10 nM. Only two		
cc	other combinations (of eleven tested) were found to have higher		
cc	values. This suggests that phage bearing scfv fragments having		
cc	weak affinities can be selected with antigen, probably due to the		
cc	avidity of the multiple antibody heads on the phage.		
cc	See also AAR21260-307, 309-311; AAR22450, 565-581.		
xx			
so	Sequence	113 AA;	
xx			
xx	Query Match	100.0%;	Score 29;
xx	Best Local Similarity	100.0%;	DB 13;
xx	Matches	5;	Conservative 0;
xx		Mismatches	0;
xx		Indels	0;
xx		Gaps	0;
oy	1 SYGVH 5		
xx			
db	31 sygvh 35		
xx			
xx	RESULT 7		
xx	AAW04595		
xx	ID AAW04595 standard; Protein; 115 AA.		
xx	AAW04595;		
xx	12-AUG-1997 (first entry)		
xx			
de	Anti-DNA antibody 11f8 heavy chain variable region.		
xx			
xx	Heavy Chain; variable region; anti-DNA; monoclonal; antibody;		
xx	MAB 11f8; haplpin; diagnosis; inflammatory glomerulonephritis;		
xx	systemic lupus erythematosus; screening; treatment; prevention;		
xx	SLE; disease.		
xx			
os	Mus spp.		
xx			
xx	Key	Location/Qualifiers	
xx			

```

FT      Region                               1..22  

FT      /label= framework_I  

FT      Region                               23..27  

FT      /label= CDR_I  

FT      Region                               28..41  

FT      /label= framework_II  

FT      Region                               42..58  

FT      /label= CDR_II  

FT      Region                               59..90  

FT      /label= framework_III  

FT      Region                               91..102  

FT      /label= CDR_III  

FT      Region                               103..115  

FT      /label= J_region  

XX  

XX      WO9636361-A1.  

XX  

XX      PD      21-NOV-1996.  

XX  

XX      PF      16-MAY-1996;    96WO-US07113.  

XX  

XX      PR      18-MAY-1995;    95US-0443540.  

XX  

XX      PA      (UNMI ) UNIV MICHIGAN.  

XX  

XX      PI      Glick GD, Swanson PC;  

XX  

XX      DR      WPI: 1997-011854/01.  

XX      N-PSDB: AAT43740.  

XX  

XX      PT      Anti-DNA antibody which specifically binds DNA hairpin - useful to  

XX      develop prods. for diagnosis and treatment of disorders, e.g.  

XX      glomerulonephritis or systemic lupus erythematosus  

XX  

XX      PS      Example: Fig 7; 102pp; English.  

XX  

CC      The present sequence is the heavy chain variable region of the  

CC      anti-DNA monoclonal antibody (Mab) 11f8, which has a high affinity  

CC      for single stranded DNA, low or no affinity for double stranded DNA  

CC      and specifically binds a DNA hairpin. The Mab can be used diagnose  

CC      disorders associated with the pathological complexation of DNA,  

CC      e.g., inflammatory glomerulonephritis and systemic lupus  

CC      erythematosus. It can also be used to generate reagents to screen  

CC      for pharmaceutical agents, and treat and/or prevent an above  

CC      disorder.  

CC      Calf thymus DNA was used to immunise a MRJ-1pr mouse, spleen cells  

CC      from which were then fused with SP2/O myeloma cells to give  

CC      hybridomas producing the anti-DNA Mab. 1f8 was found to react  

CC      strongly with single stranded DNA and poly(dT).  

CC      xx  

SQ      Sequence      115 AA:  
  

Query Match          100.0%; Score 29; DB 18; Length 115;  

Best Local Similarity 100.0%; Pred. No. 82;  

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  

QY      1 SYGVH 5  

       |||||  

Db      23 sygvh 27  
  

RESULT      8  

ID      AAR07322 standard; protein; 116 AA.  

XX      AAR07322;  

AC      AAR07322;  

XX  

DT      22-JAN-1991 (first entry)  

XX  

DE      VH domain of antibody D against tumour-associated antigens.  

XX  

KW      Tumour-associated antigen; murine monoclonal antibody D;
```

KM vibrio cholerae neuraminidase-sensitive epitope; glioma;
 KM meningioma; neurilemmoma; neuroblastoma; ganglioblastoma;
 KM ganglioneuroma; diagnosis.
 XX
 OS Mus musculus.
 XX
 PN EP88914-A.
 XX
 PD 26-SEP-1990.
 XX
 PF 21-MAR-1990; 90EP-0105322.
 XX
 PR 24-MAR-1989; 89DE-3909799.
 XX
 PA (BEHW) BEHRINGER AG.
 XX
 PI Bosslet K, Seemann G, Sedlacek HH;
 XX
 DR WPI; 1990-291873/39.
 DR N-PSDB; AAQ06229.
 XX
 PT Monoclonal antibodies to tumour associated antigens - used for
 PT diagnosis of malignant tumours etc.
 PS Disclosure; Page 14; 18pp; German.
 XX
 PS Antibody D is produced as described in EP-141079 and binds to
 CC a vibrio cholerae neuraminidase-sensitive epitope on ganglioside GD2
 CC which occurs in gliomas, meningiomas, neurilemmomas,
 CC neuroblastomas, ganglioblastomas and ganglioneuromas.
 CC They are useful in tumour diagnosis and therapy.
 CC See also AAQ06230 for VK of Mab D, AAQ07312-15 for Mab A and B and
 CC AAQ06227-28 for Mab C.
 XX
 SQ Sequence 116 AA;
 QY
 Db 1 SYGVH 5
 11111
 26 sygvh 30
 OY
 Db 1 SYGVH 5
 11111
 26 sygvh 30
 RESULT 9
 AAR28287
 ID AAR28287 standard; Protein; 116 AA.
 XX
 AC AAR28287;
 XX
 DT 04-APR-1993 (first entry)
 XX
 DE Sequence of the VH region of the Kabat human heavy chain
 DE subgroup II of anti-mucin monoclonal antibody.
 XX
 KM Antibody; PCR primer; variable heavy chain.
 XX
 OS Homo sapiens.
 XX
 OS
 FH Key
 FT 26..30 Location/Qualifiers
 FT /label= CDR 1
 FT 44..60
 FT /label= CDR 2
 FT 93..101
 FT Region /label= CDR 3
 XX
 XX WO9218534-A.
 PN
 XX
 PD 29-OCT-1992.
 XX

PF 23-APR-1992; 92WO-GB00746.
 XX
 XX
 PR 23-APR-1991; 91GB-0008652.
 XX
 PA (ANTI-) ANTISOMA LTD.
 XX
 PI Courtenay-luck NS;
 XX
 DR WPI; 1992-382045/46.
 DR N-PSDB; AAQ30192.
 XX
 XX
 PT New peptide EPP (Glu-Pro-Thr) - selectively binds mycin
 PT expressed by epithelial tumours, used for guiding toxins or
 PT labels to tumours expressing mucin
 XX
 PS Example; Table 2, Page 28; 50pp; English.
 XX
 CC Clone B is a lymphoblastoid cell line (secreting antibody directed
 CC against a tumour-associated mucin mol.) derived from the EBV-
 CC transforming and cloning of a patient's peripheral blood B-cells.
 CC After DNA isolation, the polymerase chain reaction (PCR) was
 CC employed, using oligonucleotide primers specific for the variable
 CC light and heavy chains of immunoglobulins (see AAQ30065, AAQ30066).
 CC From this mini-prep the VH gene of the human antibody, which is
 CC designated as clone-B was isolated. Sequencing of the gene encoding the
 CC VH region of clone B provided a sequence consistent with the Kabat
 CC human heavy chain subgroup II of anti-mucin monoclonal antibody.
 XX
 SQ Sequence 116 AA;
 QY
 Db 1 SYGVH 5
 11111
 26 sygvh 30
 OY
 Db 1 SYGVH 5
 11111
 26 sygvh 30
 RESULT 10
 AAW14490
 ID AAW14490 standard; Protein; 116 AA.
 XX
 AC AAW14490;
 XX
 DT 28-JAN-1997 (first entry)
 XX
 DE Monoclonal antibody D VH.
 XX
 KM heavy; light chain; monoclonal antibody; antigen 3; marker; melanoma;
 KM permanent human tumour cell line; tumour-associated antigen; epitope;
 KM gastrointestinal tumour; pancreatic carcinoma; diagnostic; therapeutic;
 KM antigen 11; Vibrio cholera; neuraminidase-resistant; ganglioside GD2.
 XX
 OS Synthetic.
 XX
 PN EP727436-A1.
 PN
 PD 21-AUG-1996.
 XX
 PF 21-MAR-1990; 90EP-0105322.
 XX
 PR 24-MAR-1989; 89DE-3909799.
 XX
 PA (BEHW) BEHRINGER AG.
 XX
 PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;
 XX
 DR WPI; 1996-372836/38.
 DR N-PSDB; AAT63507.
 XX
 PT Monoclonal antibody to tumour-associated antigen - useful as

PT gastrointestinal tumour marker
 XX
 PS Disclosure; Page 14; 19pp; German.
 CC
 XX
 CC AAM14490-91 are the heavy and light chains (respectively) of monoclonal
 CC antibody (Mab) D. Mab D recognises Vibrio cholera
 CC neuraminidase-resistant epitope of ganglioside GD2, from a human melanoma
 CC cell line. Mabs A, B and C (see AAM14484-89) are mentioned in the
 CC specification, but are not part of the claims. Mabs A and B recognise
 CC antigens 3 and 11 resp., of a permanent human tumour cell line. Mab C
 CC also recognises an epitope of a tumour-associated antigen. These antigens
 CC occur at high concns. in the serum of patients with gastrointestinal
 CC tumours, e.g. pancreatic carcinoma, and are thus useful as tumour markers
 CC for diagnostic or therapeutic purposes.
 CC
 XX
 SQ Sequence 116 AA;
 OY 1 SYGVH 5
 DB 26 sygvh 30
 Query Match 100.0%; Score 29; DB 17; Length 116;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SYGVH 5
 DB 26 sygvh 30
 RESULT 11
 AAR9877
 ID AAR9877 standard; Protein: 116 AA.
 AC AAR9877;
 XX 28-JAN-1997 (first entry)
 XX
 DE Monoclonal antibody D VH.
 XX
 KW Monoclonal antibody; Mab; epitope: tumour-associated antigen;
 KW marker: antigen.
 XX
 OS Synthetic.
 XX
 PN EP727435-A1.
 PD 21-AUG-1996.
 XX
 PF 21-MAR-1990; 90EP-0105322.
 XX
 PR 24-MAR-1989; 89DE-3909799.
 XX
 PA (BEHW) BEHRINGWERKE AG.
 XX
 PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;
 XX
 DR WPI: 1996-372835/38.
 DR N-PSDB; AAT3665.
 XX
 PT Monoclonal antibody to tumour-associated antigen - useful as
 PT gastrointestinal tumour marker
 XX
 PS Disclosure; Page 14; 19pp; German.
 XX
 CC Mab C (AAT36659-T36660) is a monoclonal antibody that recognises an
 CC epitope of a tumour-associated antigen occurring at high concn. in
 CC the serum of patients with gastrointestinal tumours, e.g. pancreatic
 CC carcinoma, and is thus useful as a tumour marker for diagnostic or
 CC therapeutic purposes.
 CC Mabs A, B and D are mentioned in the specification, but are not
 CC part of the claims.
 CC Mab A (AAT36661-T36662) recognises antigen 3 of permanent human
 CC tumour cell line.
 CC Mab B (AAT36663-T36664) recognises antigen 11 of permanent human
 CC tumour cell line.

CC Mab D (AAT36665-T36666) recognises a Vibrio cholera neuraminidase-
 CC resistant epitope of ganglioside GD2, from a human melanoma cell
 CC line.
 CC
 XX
 SQ Sequence 116 AA;
 OY 1 SYGVH 5
 DB 26 sygvh 30
 Query Match 100.0%; Score 29; DB 17; Length 116;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SYGVH 5
 DB 26 sygvh 30
 RESULT 12
 AAR25728
 ID AAR25728 standard; Protein: 119 AA.
 AC AAR25728;
 XX 13-JAN-1993 (first entry)
 XX
 DE Humanised VH region of the mouse mik-beta1 antibody.
 XX
 KW Murine; immunoglobulin; CDR; non immunogenic; interleukin-2;
 KW IL-2; heavy chain; variable region; framework; p75; human; lay.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FH Region 31..35
 FT /note= "CDR"
 FT Region 50..65
 FT /note= "CDR"
 FT Region 98..108
 FT /note= "CDR"
 FT Misc-difference 1
 FT /note= "mutated residue"
 FT Misc-difference 29
 FT /note= "mutated residue"
 FT Misc-difference 30
 FT /note= "mutated residue"
 FT Misc-difference 49
 FT /note= "mutated residue"
 FT Misc-difference 72
 FT /note= "mutated residue"
 FT Misc-difference 73
 FT /note= "mutated residue"
 FT Misc-difference 84
 FT /note= "mutated residue"
 FT Misc-difference 89
 FT /note= "mutated residue"
 FT Misc-difference 90
 FT /note= "mutated residue"
 XX
 PN W09211018-A.
 PD 09-JUL-1992.
 XX
 PF 19-DEC-1991; 91WO-US09711.
 XX
 PR 19-DEC-1990; 90US-0634278.
 XX
 PA (PROT-) PROTEIN DESIGN LABS INC.
 XX
 PI Co MS, Coelingh KL, Landolfi NF, Queen CL, Schneider WP;
 XX
 DR WPI: 1992-249842/30.
 XX
 PT New immunoglobulin(s) having murine CDRs in human framework
 PT regions - have lower antigenicity; useful for treating e.g. HSV,

PT CMV, T-cell disorders, myeloid disorders and auto-immune
PT conditions
XX
PS Claim 6; Fig 10B; 141pp; English.
XX
CC The sequence shows the humanised mature heavy chain variable
CC region of the mouse mAb-Detel antibody. Murine CDRs were used
CC in a human lay framework to produce a pure humanised immunoglobulin
CC (1g) which is capable of binding to the p75 chain of the
CC interleukin-2 receptor. The Ig is non immunogenic, due to the
CC human framework, and has a strong affinity for its predetermined
CC antigen. They can be produced in large quantities via recombinant
CC DNA and monoclonal antibody technology. The humanised Igs may be
CC used alone or in combination with chemotherapeutic agents such as
CC non-steroidal anti-inflammatory drugs or immunosuppressants.
CC See also AAR25721-32.
XX
SQ Sequence 119 AA;

Query Match 100.0%; Score 29; DB 13; Length 119;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
| | | | |
Db 31 sygvh 35

RESULT 13
AA49218
ID AAY49218 standard; Protein; 119 AA.
XX
XX AAY49218;
AC
XX 07-FEB-2000 (first entry)
DT
XX
XX Heavy chain variable region consensus sequence.
DE
XX
XX Monoclonal antibody; MAb: 1A7; GD2; immune response; melanoma;
KM neuroblastoma; glioma; soft tissue carcinoma; small cell carcinoma;
KM tumor-associated antigen.
XX
XX Synthetic.
OS
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 99..102 "residues Xaa are unspecified"
FT /note="residues Xaa are unspecified"
XX
XX US5977316-A.
PN
XX
XX 02-NOV-1999.
PD
XX
XX 16-JAN-1996; 96US-0591196.
PE
XX
XX 17-JAN-1995; 95US-0372676.
PR
XX
XX (KENT) UNIV KENTUCKY.
PA
XX
XX Foon KA, Chatterjee SK, Chatterjee M;
PI
XX
XX WPI; 1999-619711/53.
DR
XX
XX Monoclonal antibody 1A7 which elicits an anti-GD2 immunological
PT response, useful for the development of products for the detection and
PT treatment of cancers -
XX
XX Disclosure; Fig 3C; 74pp; English.
PS
XX
XX The invention provides a monoclonal antibody (MAb) designated 1A7, which
CC elicits an anti-GD2 (tumor-associated antigen) immunological response in
CC humans. MAb 1A7 has defined light and heavy chain variable region
CC sequences. The MAb 1A7 and polypeptides can be used for eliciting an

CC anti-GD2 immune response. The polypeptides can also be used for detecting
CC or purifying anti-GD2 antibody. The products can be used for treating GD2
CC -associated diseases, e.g. melanoma, neuroblastoma, glioma, soft tissue
CC carcinoma, and small cell carcinoma. They can be used for palliating the
CC disease or for reducing the risk of recurrence. The present sequence
CC represents the heavy chain variable region consensus sequence.
XX
SQ Sequence 119 AA;

Query Match 100.0%; Score 29; DB 20; Length 119;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
| | | | |
Db 31 sygvh 35

RESULT 14
AAY39808
ID AAY39808 standard; peptide; 119 AA.
XX
XX AAY39808;
AC
XX 30-NOV-1999 (first entry)
DT
XX
XX K5B8 antibody heavy chain.
DE
XX
XX Antibody; K5H8 antibody; TRL 6 antibody; activation domain; HIV-1; Tat;
KM replication inhibitor; diagnosis; HIV infection; heavy chain.
KM
XX
XX Synthetic.
OS
XX
XX WO9945959-A1.
FN
XX
XX 16-SEP-1999.
PD
XX
XX 11-MAR-1999; 99WO-US05262.
PE
XX
XX 13-MAR-1998; 98US-0077845.
PR
XX
XX (DAND) DANA FARBER CANCER INST INC.
PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
PA (CHIR) CHIRON CORP.
XX
XX Marasco W, Lavecchio J, Mhashilkar AM, Ramstedt U, Ring D;
PI Eberhardt B, Porter-Brooks J;
XX
XX WPI; 1999-561622/47.
DR
XX
XX Novel humanized antibodies, use as, e.g. anti-HIV-1 agents -
PT
XX
XX Claim 1; Fig 7; 64pp; English.
PS
XX
XX This sequence represents the heavy chain from the K5H8 antibody. The
CC invention relates to a humanized antibody framework motif that is
CC selected from a human library based upon comparison to a murine antibody,
CC and the heavy chain is encoded by the VH gene of K5B8, and the light
CC chain is encoded by the VL gene of TRL 6. The humanized antibodies can be
CC used in a variety of ways, and may be expressed intracellularly. For
CC example, they can be directed against the activation domain of HIV-1, and
CC so inhibit HIV-1 replication. They can also be used extracellularly to
CC target Tat, and, when combined with a detectable label, be used to
CC measure levels of Tat for diagnosis or prognosis, and to diagnose
CC susceptibility to HIV infection. In order to minimize risk of evoking an
CC immune response, murine anti-Tat antibodies are humanized. However, prior
CC art methods have resulted in humanized antibodies that are not as
CC effective as the original antibody. Frequently, the humanized antibody
CC will retain some murine amino acids. It would be desirable to have a
CC framework motif that produces an antibody having a protective efficiency
CC comparable to the murine antibody. This need is met by the humanized
CC antibodies. By retaining no murine amino acids in the humanized

CC framework, a more effective antibody against HIV is produced, compared to
CC prior art methods.
XX
SQ Sequence 119 AA;

OY 1 SYGVH 5
Db 31 sygvh 35

Query Match 100.0%; Score 29; DB 20; Length 119;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: August 20, 2002, 13:15:09
Job time: 103 sec

OY 1 SYGVH 5
Db 31 sygvh 35

RESULT 15
AAB69660
ID AAB69660 standard; Protein: 119 AA.
XX
AC AAB69660;
XX
DT 30-APR-2001 (first entry)
XX
DE Human lay antibody heavy chain SEQ ID NO: 37.
XX
KW Humanised immunoglobulin; mouse; human; antibody; heavy chain; diabetes;
KW light chain; graft versus host disease; transplant; autoimmune disease;
KW multiple sclerosis; rheumatoid arthritis; systemic lupus erythematosus;
KW myasthenia gravis; herpes infection; myeloid leukaemia; CMV infection.
XX
OS Homo sapiens.
XX
PN US6180370-B1.
XX
PD 30-JAN-2001.
XX
PE 07-JUN-1995; 95US-0484537.
XX
PR 28-DEC-1988; 88US-0290975.
PR 13-FEB-1989; 89US-0310252.
PR 28-SEP-1990; 90US-0590274.
PR 19-DEC-1990; 90US-0634278.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
PI Queen CL, Selick HE;
PI
DR WPI: 2001-190856/19.
XX
PT Producing humanized immunoglobulin, involves producing a cell
PT containing DNA segments encoding humanized heavy and light chain
PT variable regions, and expressing the DNA segments in the cell -
XX
PS Example 5; Fig 26; 145pp; English.
XX
CC The present invention describes a method of producing humanised
CC immunoglobulins involving expressing in a cell a nucleic acid encoding a
CC humanised version of an immunoglobulin. This is obtained by comparing a
CC donor and human immunoglobulin and producing a combined antibody which
CC contains part of each. These are useful in the treatment of
CC graft-versus-host disease, transplant rejection, autoimmune diseases such
CC as diabetes, rheumatoid arthritis, myasthenia gravis, multiple sclerosis
CC and systemic lupus erythematosus, herpes infections, CMV virus infections
CC and myeloid leukaemia. The present sequence is an antibody used to
CC demonstrate the method of the invention.
XX
SQ Sequence 119 AA;

Query Match 100.0%; Score 29; DB 22; Length 119;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:13:26 ; Search time 46.28 Seconds

(Without alignments)
10.381 Million cell updates/sec

Title: US-09-824-286-3_COPY_28_32

Perfect score: 29

Sequence: 1 SYGVH 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	100.0	76	2 AH2120	hypothetical prote
2	29	100.0	90	2 A49042	Ig heavy chain V r
3	29	100.0	95	2 S17605	Ig heavy chain V r
4	29	100.0	97	2 S55372	Ig heavy chain V r
5	29	100.0	98	2 S21812	Ig heavy chain V r
6	29	100.0	100	2 A25913	Ig heavy chain pre
7	29	100.0	100	2 S14490	Ig heavy chain V r
8	29	100.0	101	2 S03466	Ig heavy chain V r
9	29	100.0	101	2 PH1021	Ig heavy chain V r
10	29	100.0	105	2 PH1020	Ig heavy chain V r
11	29	100.0	106	2 S26322	Ig heavy chain V r
12	29	100.0	106	2 S14489	Ig heavy chain V r
13	29	100.0	107	2 S14492	Ig heavy chain V r
14	29	100.0	107	2 S14491	Ig heavy chain V r
15	29	100.0	108	2 S07331	Ig heavy chain V r
16	29	100.0	109	2 PH1023	Ig heavy chain V r
17	29	100.0	109	2 PH1025	Ig heavy chain V r
18	29	100.0	111	2 PH1019	Ig heavy chain V r
19	29	100.0	112	2 PH1022	Ig heavy chain V r
20	29	100.0	112	2 S11098	Ig heavy chain V r
21	29	100.0	112	2 S11100	Ig heavy chain V r
22	29	100.0	112	2 S11108	Ig heavy chain V r
23	29	100.0	113	2 S11101	Ig heavy chain V r
24	29	100.0	114	2 S26321	Ig heavy chain V r
25	29	100.0	114	2 S11099	Ig heavy chain V r
26	29	100.0	114	2 S11104	Ig heavy chain V r
27	29	100.0	114	2 S11106	Ig heavy chain V r
28	29	100.0	114	2 S11105	Ig heavy chain V r
29	29	100.0	115	2 S26470	Ig heavy chain V r

30	29	100.0	115	2 S11107	Ig heavy chain V r
31	29	100.0	116	1 G1MS10	Ig heavy chain pre
32	29	100.0	116	2 A33932	Ig mu chain precu
33	29	100.0	116	2 S11102	Ig heavy chain V r
34	29	100.0	117	2 S10111	Ig heavy chain V r
35	29	100.0	127	2 B31807	Ig heavy chain V r
36	29	100.0	135	2 S31913	Ig gamma-2A chain
37	29	100.0	141	2 S52446	Ig heavy chain V r
38	29	100.0	144	2 S11244	Ig gamma-2a chain
39	29	100.0	240	2 D89916	dihydrodipicolinat
40	29	100.0	285	2 S29306	poly(3-hydroxylaika
41	29	100.0	288	2 E89903	hypothetical prote
42	29	100.0	305	2 T49720	hypothetical prote
43	29	100.0	397	2 F90182	hypothetical prote
44	29	100.0	587	2 AE0487	probable hemolysin
45	29	100.0	657	2 T12969	hypothetical prote

ALIGNMENTS

RESULT 1
AH2120
hypothetical protein asl2519 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence-revision 14-Dec-2001 #text-change 11-Jan-2002
C:Accession: AH2120
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iridu
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-76 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA874218.1; PID:g17131611; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: asl2519

Query Match 100.0%; Score 29; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYGVH 5
|||||
Db 31 SYGVH 35

RESULT 2
A49042
Ig heavy chain V region (anti-DNA) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Dec-1993 #sequence-revision 18-Nov-1994 #text-change 23-Jul-1999
C:Accession: A49042
R:Taki, S.; Hirose, S.; Kinoshita, K.; Nishimura, H.; Shimamura, T.; Hamuro, J.; Shlr
Eur. J. Immunol. 22, 987-992, 1992
A:Title: Somatic mutation of anti-DNA antibody clonally related to germ-line enco
A:Reference number: A49042; MUID:92201320
A:Accession: A49042
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-90 <TAK>
A:Cross-references: GB:S92270; NID:g247954; PIDN:AAB21904.1; PID:g247955
A:Experimental source: NZB, NZW, liver
C:Note: sequence extracted from NCBI backbone (NCBI:92270, NCBI:92271)
C:Superfamily: immunoglobulin V region; immunoglobulin homology
F:8-90/Domain: immunoglobulin homology <IMK>

Query Match 100.0%; Score 29; DB 2; Length 90;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SYGVH 5
|||||
Db 24 SYGVH 28

RESULT 3
S17605
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 06-Jun-1997
C:Accession: S17605
R:Clackson, T.; Hoogenboom, H.R.; Griffiths, A.D.; Winter, G.
Nature 352, 624-628, 1991
A:Title: Making antibody fragments using phage display libraries.
A:Reference number: S17230; MUID:91326098
A:Accession: S17605
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-95 <CLa>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:7-89/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 24 SYGVH 28

RESULT 4
S55372
Ig heavy chain V region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 12-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 23-Jul-1999
C:Accession: S55372
R:Well, M.; Grimsrud, K.; Simon, D.; Lefranc, G.; Pau, B.; Lefranc, M.P.
submitted to the EMBL Data Library, November 1990
A:Description: Sequence of the 4C1 antibody heavy and light chain variable regions.
A:Reference number: S55366
A:Accession: S55372
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-97 <MEI>
A:Cross-references: EMBL:X56177; NID:9854282; PIDN:CAA39638.1; PID:9854283
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:7-89/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 23 SYGVH 27

RESULT 5
S21812
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 16-Aug-1996
C:Accession: S21812
R:Mocklat, R.

submitted to the EMBL Data Library, March 1991
A:Reference number: S21812
A:Accession: S21812
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-98 <MOC>
A:Cross-references: EMBL:X58634
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:7-89/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 23 SYGVH 27

RESULT 6
A25913
Ig heavy chain precursor V region (BFL2) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 21-Jan-2000
C:Accession: A25913
R:Lawler, A.M.; Lin, P.S.; Gearhart, P.J.
Proc. Natl. Acad. Sci. U.S.A. 84, 2454-2458, 1987
A:Title: Adult B-cell repertoire is biased toward two heavy-chain variable-region genes
A:Reference number: A94148; MUID:87175692
A:Accession: A25913
A:Molecule type: DNA
A:Residues: 1-100 <LA>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:19-100/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 35 SYGVH 39

RESULT 7
S14490
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C:Accession: S14490
R:Chen, O.; Stenzel-Poore, M.; Rittenberg, M.B.
submitted to the EMBL Data Library, March 1991
A:Description: Natural polyclonal antibodies differ from Ag-induced antibodies in V
A:Reference number: S14484
A:Accession: S14490
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-100 <CHR>
A:Cross-references: EMBL:X58647; NID:951283; PIDN:CAA41504.1; PID:951284
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5

Db 31 SYGVH 35

RESULT 8

S03466
Ig heavy chain V region (Vgb) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 16-Aug-1996
C:Accession: S03466
R:Rech, M.G.; Jackson, S.; Alt, F.W.
EMBO J. 5, 2131-2138, 1986
A:Title: V(H)DJ(H) formation and DJ(H) replacement during pre-B differentiation: non-rat
A:Reference number: S03466; MUID:87053850
A:Accession: S03466
A:Molecule type: DNA
A:Residues: 1-101 <RET>
A:Cross-references: EMBL:X04348
A:Note: the authors translated the codon ATC for residue 113 as Ser
A:Note: this sequence was determined from the nonfunctional differentiated gene
C:Genetics:
A:Map position: 12
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:18-100/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYGVH 5
Db 34 SYGVH 38

RESULT 9

PH1021
Ig heavy chain V region (clone 165.52) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C:Accession: PH1021
R:Fillman, D.M.; Jou, N.T.; Hall, R.J.; Marion, T.N.
J. Exp. Med. 176, 761-779, 1992
A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective B
A:Reference number: PH0971; MUID:92381444
A:Accession: PH1021
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-101 <TIL>
A:Experimental source: B cell, strain [NZB x NZW]F1
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:7-89/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYGVH 5
Db 23 SYGVH 27

RESULT 10

PH1020
Ig heavy chain V region (clone 165.41) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C:Accession: PH1020
R:Fillman, D.M.; Jou, N.T.; Hall, R.J.; Marion, T.N.
J. Exp. Med. 176, 761-779, 1992

A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective
A:Reference number: PH0971; MUID:92381444

A:Accession: PH1020
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-105 <TIL>
A:Experimental source: B cell, strain [NZB x NZW]F1
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:14-95/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYGVH 5
Db 29 SYGVH 33

RESULT 11

S26322
Ig heavy chain V region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1998 #sequence_revision 19-Mar-1998 #text_change 21-Jan-2000
C:Accession: S26322
R:Stark, S.E.; Caton, A.J.
J. Exp. Med. 174, 613-624, 1991
A:Title: Antibodies that are specific for a single amino acid interchange in a proteol
A:Reference number: S26309; MUID:91341421
A:Accession: S26322
A:Molecule type: mRNA
A:Residues: 1-106 <STA>
A:Cross-references: EMBL:X59182
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:9-91/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYGVH 5
Db 25 SYGVH 29

RESULT 12

S14489
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C:Accession: S14489
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.
submitted to the EMBL data library, March 1991
A:Description: Natural polyclonal antibodies differ from Ag-induced antibodies in V
A:Reference number: S14484
A:Accession: S14489
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-106 <CHE>
A:Cross-references: EMBL:X58646; NID:951281; PTDN:CAA41503.1; PTD:951282
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 31 SYGVH 35

RESULT 13

SI4492
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C:Accession: SI4492
R:Chen, O.; Stenzel-Poore, M.; Rittenberg, M.B.
submitted to the EMBL Data Library, March 1991
A:Description: Natural polyclonal antibodies differ from Ag-induced antibodies in VH C
A:Reference number: SI4484
A:Accession: SI4492
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-107 <CHE>
A:Cross-references: EMBL:X58649; NID:g51287; PIDN:CAA41506.1; PID:g51288
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 31 SYGVH 35

RESULT 14

SI4491
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C:Accession: SI4491
R:Chen, O.; Stenzel-Poore, M.; Rittenberg, M.B.
submitted to the EMBL Data Library, March 1991
A:Description: Natural polyclonal antibodies differ from Ag-induced antibodies in VH C
A:Reference number: SI4484
A:Accession: SI4491
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-107 <CHE>
A:Cross-references: EMBL:X58648; NID:g51285; PIDN:CAA41505.1; PID:g51286
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 31 SYGVH 35

RESULT 15

S07331
Ig heavy chain V region (clone NO2-12.4) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Feb-1997 #sequence_revision 18-Jul-1997 #text_change 21-Jan-2000
C:Accession: S07331
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467

A:Accession: S07331
A:Molecule type: mRNA
A:Residues: 1-108 <KAA>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 31 SYGVH 35

Search completed: August 20, 2002, 13:16:02
Job time: 156 sec

2002-08-20 13:05:56

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:13:31 ; Search time 21.48 Seconds

(without alignments)
9.013 Million cell updates/sec

Title: US-09-824-286-3_COPY_28_32
Perfect score: 29
Sequence: 1 SYGVH 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	116	1 HV45_MOUSE	P01821 mus musculus
2	29	100.0	220	1 DABP_STRAU	Q92111 staphylococ
3	29	100.0	1856	1 MGA_HUMAN	Q43421 homo sapien
4	29	96.6	187	1 PAAD_HELPJ	Q92363 helicobacte
5	28	96.6	187	1 PAAD_HELPJ	Q26011 helicobacte
6	28	96.6	840	1 MUTS_LACTA	Q9CD99 lactococcus
7	28	96.6	844	1 HEXA_STRPN	P01564 streptococc
8	28	96.6	851	1 MUTS_STRPN	Q99X18 streptococc
9	28	96.6	858	1 MUTS_BACSU	P49849 bacillus su
10	26	89.7	89	1 RS15_MYCLE	Q32967 mycobacteri
11	26	89.7	139	1 Y965_MYCTU	P71545 mycobacteri
12	26	89.7	148	1 CU08_LOCM1	P11734 locusta mig
13	26	89.7	185	1 3MGR_RHIME	Q921C1 ribzobium m
14	26	89.7	195	1 GRP3_DAUCA	P37705 daucus caro
15	26	89.7	198	1 YN00_YEAST	P53846 saccharomyc
16	26	89.7	211	1 THIE_ECOLI	P30137 escherichia
17	26	89.7	211	1 THIE_SALTY	Q91918 salmonella
18	26	89.7	412	1 PT56_YEAST	P25270 saccharomyc
19	26	89.7	435	1 YFS3_CAEEL	P49191 caenorhabdi
20	26	89.7	452	1 HOS2_YEAST	P53056 saccharomyc
21	26	89.7	474	1 M3010_HABIN	P73010 haemophilus
22	26	89.7	524	1 SYK_CERSY	Q74059 cenarchaeum
23	26	89.7	583	1 NU2M_NEUCR	Q35140 neurospora
24	26	89.7	658	1 CTR2_HUMAN	P52569 homo sapien
25	26	89.7	676	1 TKT7_CRAPL	Q42677 cratereostilg
26	26	89.7	679	1 TKTA_CRAPL	Q42675 cratereostilg
27	26	89.7	704	1 G1GB_YEAST	P32775 saccharomyc
28	26	89.7	736	1 ALD_MOUSE	P48410 mus musculu
29	26	89.7	795	1 SYFB_VIRCH	Q9SK66 vibrio chol
30	26	89.7	803	1 RIR1_CRYPV	Q61065 cryptospori
31	26	89.7	804	1 RIR1_PLAFG	P50647 plasmodium
32	26	89.7	804	1 RIR1_PLAFG	P50647 plasmodium
33	26	89.7	806	1 RIR1_PLAF4	P50648 plasmodium

34	26	89.7	835	1 UREA_SCHPO	Q00084 schizosacch
35	26	89.7	891	1 MUTS_RICPR	Q924m1 rickettsia
36	26	89.7	1051	1 YC94_HUMAN	Q924m2 homo sapien
37	26	89.7	1062	1 YAL3_HUMAN	Q924m3 homo sapien
38	26	89.7	1465	1 MTW2_HUMAN	P54296 homo sapien
39	26	89.7	1593	1 TYCC_BACBR	P58397 homo sapien
40	26	89.7	6486	1 TYCC_BACBR	Q30409 b tyrocidin
41	25	86.2	113	1 MTH_CARMA	Q27225 carcinus ma
42	25	86.2	142	1 YN8P_YEAST	P53737 saccharomyc
43	25	86.2	167	1 YBL7_SCHPO	Q9YU74 schizosacch
44	25	86.2	208	1 VEP_NPVPD	Q10318 oryza pseu
45	25	86.2	220	1 YADF_ECOLI	P36857 escherichia

ALIGNMENTS

RESULT 1					
HV45_MOUSE	STANDARD;	PRT;	116 AA.		
ID HV45_MOUSE					
PI01821:					
DT 21-JUL-1986 (Rel. 01, Created)					
DT 21-JUL-1986 (Rel. 01, Last sequence update)					
DT 15-JUL-1999 (Rel. 38, Last annotation update)					
DE Ig heavy chain V region MC101 precursor.					
OS Mus musculus (Mouse).					
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.					
OX NCBI_TaxID=10090;					
RN [1]					
RP SEQUENCE FROM N.A.					
RX MEDLINE=82075900; PubMed=6273429;					
RA Kataoka T., Nikaido T., Miyata T., Moriwa K., Honjo T.;					
RT "The nucleotide sequences of rearranged and germline immunoglobulin					
RT VH genes of a mouse myeloma MC101 and evolution of VH genes in					
RT mouse."					
RL J. Biol. Chem. 257:277-285(1982).					
CC -----					
CC This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC the European Bioinformatics Institute. There are no restrictions on its					
CC use by non-profit institutions as long as its content is in no way					
CC modified and this statement is not removed. Usage by and for commercial					
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/					
CC or send an email to license@isb-sib.ch).					
CC -----					
DR EMBL; J00502; AAA38515.1; -					
DR PIR; A02096; GIMS10.					
DR InterPro; IPR003006; Ig_MHC.					
DR InterPro; IPR003596; Ig_V.					
DR Pfam; PF00047; Ig; 1.					
DR SMART; SM00406; IGV; 1.					
KW Immunoglobulin V region; Signal.					
FT SIGNAL	1	19			
FT CHAIN	20	116			
FT NON_TER	116	116			
SO SEQUENCE	116 AA;	12593 MW;	8079A6EE7C552B3E CRC64;		
Query Match		100.0%;	Score 29;	DB 1;	Length 116;
Best Local Similarity		100.0%;	Pred. No. 5.9;		
Matches	5;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
OY 1 SYGVH 5					
DB 50 SYGVH 54					
RESULT 2					
DABP_STAU	STANDARD;	PRT;	230 AA.		
ID DABP_STAU					
AC Q9E211;					
DT 16-OCT-2001 (Rel. 40, Created)					

```

DT      16-OCT-2001 (Rel. 40, Last sequence update)
DR      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Dihydropyridicolinate reductase (EC 1.3.1.26) (DHPR).
GN      DAPB.
OS      Staphylococcus aureus.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group;
OC      Bacillus/Staphylococcus group; Staphylococcus.
OX      NCBI_TaxID=1280;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=8325-4;
RA      Wiltshire M.D., Burnham M.K., Foster S.J.;
RT      "Identification and analysis of Staphylococcus aureus components
RT      involved during growth in serum."
RL      Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC      -1- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydropyridicolinate + NAD(P)(+) -
CC      -1- 2,3-dihydropyridicolinate + NAD(P)H.
CC      -1- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
CC      semialdehyde; second step.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC      -1- SIMILARITY: BELONGS TO THE DIHYDROPIRIDIOLINATE REDUCTASE FAMILY.
CC      -----
CC      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; AF306669; AACG2247.1; -.
DR      InterPro; IPR000846; DAPB.
DR      Pfam; PF01113; DAPB; 1.
DR      PRODOM; PD004105; DAPB; 1.
DR      PROSITE; PS01298; DAPB; DAPB; FALSE; NEG.
KW      Diaminopimelate biosynthesis; lysine biosynthesis; oxidoreductase;
KW      NADP.
SQ      SEQUENCE 230 AA; 25523 MW; 10FCF79D7AF84A93 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYGVH 5
DB 97 SYGVH 101

RESULT 3
MCA_HUMAN STANDARD; PRT; 1856 AA.
AC 043451;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Maltase-glucoamylase, intestinal [Includes: Maltase (EC 3.2.1.20)
DE (Alpha-glucosidase); Glucoamylase (EC 3.2.1.3) (Glucaan 1,4-alpha-
DE glucosidase)].
DE MGAM OR MGA OR MGAML.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Small intestine;
RX MEDLINE=98112863; PubMed=9446624;
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quarzoni A.,
RA Sterchi E.E.;
RT "Human small intestinal maltase-glucoamylase cDNA cloning. Homology
RT to sucrase-isomaltase."
RL J. Biol. Chem. 273:3076-3081(1998)

```

RP	RP	[2]	CHARACTERIZATION.	
RC	RC		TISSUE=Small Intestine mucosa.	
RA	RA		MEDLINE=89066802; PubMed=3143729;	
RA	RA		Natm H.Y., Serechi E.E., Lente M.J.;	
RT	RT		"Structure, biosynthesis, and glycosylation of human small intestinal maltase-glucoamylase."	
RT	RT		J. Biol. Chem. 263:19709-19717(1988).	
RL	RL	[3]	SULFATION	
RP	RP		MEDLINE=86082658; PubMed=3121301;	
RA	RA		Danielson E.M.;	
RT	RT		"Tyrosine sulfation, a post-translational modification of microvillar enzymes in the small intestinal enterocyte."	
RT	RT		EMBO J. 6:2891-2896(1987).	
CC	CC		-I- FUNCTION: MAY SERVE AS AN ALTERNATE PATHWAY FOR STARCH DIGESTION WHEN LUMINAL ALPHA-AMYLASE ACTIVITY IS REDUCED BECAUSE OF IMMATURITY OF MALNUTRITION. MAY PLAY A UNIQUE ROLE IN THE DIGESTION OF MALTED DIETARY OLIGOSACCHARIDES USED IN FOOD MANUFACTURING.	
CC	CC		-I- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-linked D-glucose residues with release of D-glucose.	
CC	CC		-I- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.	
CC	CC		-I- SUBUNIT: MONOMER.	
CC	CC		-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. BRUSH BORDER.	
CC	CC		-I- TISSUE SPECIFICITY: EXPRESSED IN SMALL INTESTINE, GRANULOCYTE, AND KIDNEY BUT NOT IN SALIVARY GLAND OR PANCREAS.	
CC	CC		-I- PTM: N- AND O-GLYCOSYLATED.	
CC	CC		-I- PTM: DOES NOT UNDERGO INTRACELLULAR OR EXTRACELLULAR PROTEOLYTIC CLEAVAGE.	
CC	CC		-I- PTM: SULFATED (BY SIMILARITY).	
CC	CC		-I- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.	
CC	CC		-I- SIMILARITY: CONTAINS 2 P-TYPE (TREFOLI) DOMAINS.	
CC	CC		THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).	
CC	CC		-----	
DR	DR		EMBL, AF016833; AAC39568.1; -	
DR	DR		MIM, 154360; -	
DR	DR		InterPro: IPR000322; Glyco_hydro_31.	
DR	DR		InterPro: IPR000519; P_trefol1.	
DR	DR		Pfam: PF01055; Glyco_hydro_31; 2.	
DR	DR		Pfam: PF00088; trefol1; 2.	
DR	DR		SMART, SM00018; P; 2.	
DR	DR		PROSITE, PS00129; GLYCOSYL_HYDROL_F31_1; 2.	
DR	DR		PROSITE, PS00025; P_TREFOLI; 1.	
KW	KW		Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;	
KW	KW		Glycosidase; Repeat; Signal-anchor; Sulfation.	
FT	FT		INIT MET 0	
FT	FT		DOMAIN 1 12	CYTOPLASMIC (POTENTIAL).
FT	FT		TRANSMEM 13 33	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT	FT		DOMAIN 34 1856	LUMENAL (POTENTIAL).
FT	FT		DOMAIN 37 83	SER/THR-RICH.
FT	FT		DOMAIN 88 132	P-TYPE 1.
FT	FT		DOMAIN 197 914	MALTASE.
FT	FT		DOMAIN 952 998	P-TYPE 2.
FT	FT		DOMAIN 1066 1812	GLUCOAMYLASE.
FT	FT		ACT_SITE 528 528	BY SIMILARITY.
FT	FT		ACT_SITE 1419 1419	BY SIMILARITY.
FT	FT		DISULFID 89 117	BY SIMILARITY.
FT	FT		DISULFID 100 116	BY SIMILARITY.
FT	FT		DISULFID 111 129	BY SIMILARITY.
FT	FT		DISULFID 965 982	BY SIMILARITY.
FT	FT		DISULFID 977 995	BY SIMILARITY.
FT	FT		MOD_RES 415 415	SULFATION (POTENTIAL).

FT MOD_RES 424 424 Sulfation (POTENTIAL).
 FT MOD_RES 1281 1281 Sulfation (POTENTIAL).
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 826 826 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 884 884 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 976 976 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 988 988 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1254 1254 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1322 1322 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1363 1363 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1387 1387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1602 1602 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1671 1671 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1841 1841 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1846 1846 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 1856 AA; 209571 MM; 4A1903B82E768004 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 1856;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
 |||||
 DB 1165 SYGVH 1169

RESULT 4
 ID PAAD_HELPY STANDARD; PRT; 187 AA.

AC 092JEB;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable aromatic acid decarboxylase (EC 4.1.1.-).
 GN JHP1369.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 CC -1- SIMILARITY: BELONGS TO THE POLYPRENTL P-HYDROXYBENZONATE /
 CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AEO01560; AAD06958.1; -
 CC InterPro: IPR003382; Flavoprotein.
 DR Pfam: PF02441; Flavoprotein; 1.
 DR Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
 KW

SO SEQUENCE 187 AA; 20598 MM; D9BFAD9029FA396A CRC64;

Query Match 96.6%; Score 28; DB 1; Length 187;
 Best Local Similarity 80.0%; Pred. No. 16;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
 |||||
 DB 76 SYGVH 80

RESULT 5
 ID PAAD_HELPY STANDARD; PRT; 187 AA.

AC 026011;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable aromatic acid decarboxylase (EC 4.1.1.-).
 GN HPI476.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kervavage A.R., Clayton R.A., Sutton G.C.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalne H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori.";
 RL Nature 388:539-547(1997).
 CC -1- SIMILARITY: BELONGS TO THE POLYPRENTL P-HYDROXYBENZONATE /
 CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

OY 1 SYGVH 5
 |||||
 DB 76 SYGVH 80

RESULT 6
 ID MUTS_LACLA STANDARD; PRT; 840 AA.
 AC 09CDK9;

```
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein muts.
GN MUTS OR HEXA OR IL2210.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
CC Lactococcus.
CC NCBI_TaxID=1360;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=IL1403;
CC MEDLINE=21235186; PubMed=11337471;
CC Biotin A., Wincker P., Manger S., Jallion O., Malarne K.,
CC Weissbach J., Ehrlich S.D., Sorokin A.;
CC "The complete genome sequence of the lactic acid bacterium Lactococcus
CC lactis ssp. lactis IL1403."
CC Genome Res. 11:731-753(2001).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step. This protein has a weak ATPase activity (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AE006450; AAK06308.1; -
CC InterPro: IPR000432; MUTS_C.
CC InterPro: IPR002863; MUTS_N.
CC Pfam: PF00488; MUTS_C; 1.
CC Pfam: PF01624; MUTS_N; 1.
CC ProDom: PD001263; MUTS_C; 1.
CC SMART: SM00533; MUTSd; 1.
CC PROSITE: PS00486; DNA_MISMATCH_REPAIR_2; 1.
CC DNA repair: ATP-binding; DNA-binding; Complete proteome.
CC NP_BIND 601 608 ATP (POTENTIAL).
CC SEQUENCE 840 AA; 94272 MW; 867EFE4F82616DC2 CRC64;

OY 1 SYGVH 5
DB 757 SYGII 761

Query Match 96.6%; Score 28; DB 1; Length 840;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
RT Streptococcus pneumoniae and homolog of hexA to muts of Escherichia
RT coli and Salmonella typhimurium";
RT J. Bacteriol. 170:190-196(1988).
RN [2]
RN SEQUENCE FROM N.A.
RN STRAIN=TIGRA;
RN MEDLINE=21357209; PubMed=11463916;
RN Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RN Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RN Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RN Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RN Holtzapple E., Kouri H., Wolf A.M., Ueberback T.R., Hansen C.L.,
RN McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RN Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RN Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RT Science 293:498-506(2001).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step.
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT (POSSIBLY NATURAL) IN POSITION 759.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M18729; AAA88597.1; -
CC EMBL, AE007496; -; NOT_ANNOTATED_CDS.
CC PIR: C28667; C28667.
CC TIGR: SP2076; -
CC InterPro: IPR000432; MUTS_C.
CC InterPro: IPR002863; MUTS_N.
CC Pfam: PF00488; MUTS_C; 1.
CC Pfam: PF01624; MUTS_N; 1.
CC ProDom: PD001263; MUTS_C; 1.
CC SMART: SM00533; MUTSd; 1.
CC PROSITE: PS00486; DNA_MISMATCH_REPAIR_2; 1.
CC DNA repair: ATP-binding; DNA-binding; Complete proteome.
CC NP_BIND 602 609 ATP (POTENTIAL).
CC SEQUENCE 844 AA; 94048 MW; 59C1B963D90B6EE2 CRC64;

OY 1 SYGVH 5
DB 758 SYGII 762

Query Match 96.6%; Score 28; DB 1; Length 844;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC:700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Seate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*,"
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step. This protein has a weak ATPase activity (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE006633; AK34785.1; -
DR InterPro: IPR000432; MUTS_C.
DR InterPro: IPR002863; MUTS_N.
DR Pfam: PF00488; MUTS_C; 1.
DR Pfam: PF01624; MUTS_N; 1.
DR Prodom: PD001263; MUTS_C; 1.
DR SMART: SM00534; MUTSacc; 1.
DR SMART: SM00533; MUTSD; 1.
DR PROSITE: PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT NP_BIND 602 609 ATP (POTENTIAL).
SQ SEQUENCE 851 AA; 95470 MW; 737D51C8DCEFAF0A CRC64;

Query Match 96.6%; Score 28; DB 1; Length 851;
Best Local Similarity 80.0%; Pred. No. 73;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||:|
Db 758 SYGIV 762

RESULT 9
MUTS_BACSU STANDARD: PRT; 858 AA.
ID MUTS_BACSU
AC P49849;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein mutS.
GN MUTS.
OS *Bacillus subtilis*.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Ginetti F., Perego M., Albertini A.M., Galizzi A.;
RT "Bacillus subtilis mutS mult operon: identification, nucleotide
RT sequence and mutagenesis,"
RL Microbiology 142:2021-2029(1996).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step. This protein has a weak ATPase activity (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U27343; AB19235.1; ALT_INIT.
DR EMBL: 299112; CAB1357.1; -
DR Subtilist: BG11403; mutS.
DR InterPro: IPR000432; MUTS_C.
DR InterPro: IPR002863; MUTS_N.
DR Pfam: PF00488; MUTS_C; 1.
DR Pfam: PF01624; MUTS_N; 1.
DR Prodom: PD001263; MUTS_C; 1.
DR SMART: SM00534; MUTSacc; 1.
DR SMART: SM00533; MUTSD; 1.
DR PROSITE: PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT NP_BIND 602 609 ATP (POTENTIAL).
SQ SEQUENCE 858 AA; 97590 MW; 641BDCDBCA4BC5D CRC64;

Query Match 96.6%; Score 28; DB 1; Length 858;
Best Local Similarity 80.0%; Pred. No. 74;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||:|
Db 758 SYGIV 762

RESULT 10
RS15_MYCLE STANDARD: PRT; 89 AA.
ID RS15_MYCLE
AC O32967;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S15.
GN RPSO OR M08953 OR MLCB22.28C.
OS *Mycobacterium leprae*.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; *Mycobacterium*.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN.
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Whittall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
RA Holtroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus,"
RL Nature 409:1007-1011(2001).
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE 16S RIBOSOMAL RNA BINDING
CC PROTEINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S15P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z98741; CAB11393.1; -
DR EMBL: AL583920; CAC31234.1; -

DR HSSP: P05766; 1A32.
DR Leptoma; ML0853; -;
DR InterPro: IPR000589; Ribosomal_S15.
DR Pfam: PF00312; Ribosomal_S15; 1.
DR PROSITE: PS00362; RIBOSOMAL_S15; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 89 AA: 10348 MW: 4175AC55204571EE CRC64;

Query Match 89.7%; Score 26; DB 1; Length 89;
Best Local Similarity 80.0%; Pred. NO. 23;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
111:1
Db 14 SYGLH 18

RESULT 11
RS15_MYCTU STANDARD: PRT; 89 AA.
AC O33327;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S15.
GN RPSO OR RV2785C OR MT2855 OR MTW002.50C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE 16S RIBOSOMAL RNA BINDING
CC PROTEINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S15P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AL008967; CA15580.1; -;
DR EMBL: AE007112; AAK47174.1; -;
DR HSSP: P05766; 1A32.
DR TIGR: MT2855; -;

DR TubercuList: RV2785C; -;
DR InterPro: IPR000589; Ribosomal_S15.
DR Pfam: PF00312; Ribosomal_S15; 1.
DR PROSITE: PS00362; RIBOSOMAL_S15; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 89 AA: 10475 MW: 6DD2577898F97D7 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 89;
Best Local Similarity 80.0%; Pred. NO. 23;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
111:1
Db 14 SYGLH 18

RESULT 12
Y965_MYCTU STANDARD: PRT; 139 AA.
ID Y965_MYCTU
AC P71545;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 14.5 kDa protein RV0965C.
GN RV0965C OR MT0993 OR MTCY10D7.09.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO M.TUBERCULOSIS RV2798C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z79700; CAH01986.1; -;
DR EMBL: AE006984; AAK45242.1; ALT_INIT.
DR TIGR: MT0993; -;
DR TubercuList: RV0965C; -;
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 139 AA: 14479 MW: D6DAF1B4D1A1332 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 139;
 Best Local Similarity 80.0%; Pred. No. 35;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
 |||||
 DB 99 SYGMH 103

RESULT 13

C008_LOCM1 STANDARD; PRT; 148 AA.
 AC P11734;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Cuticle protein 8 (Lm-8) (Lm-ACP 8).
 OS Locusta migratoria (Migratory locust).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
 OC Acridoidea; Acrididae; Locusta.
 NC NCBI_TaxID=7004;
 RX MEDLINE=90073593; PubMed=2590176;
 RA Klarskov K., Hoejrup P., Andersen S.O., Roepstorff P.;
 RT "Plasma-desorption mass spectrometry as an aid in protein sequence
 determination. Application of the method on a cuticular protein from
 the migratory locust (Locusta migratoria).";
 RL Biochem. J. 262:923-930(1989).
 RN [2]

SEQUENCE OF 1-56.
 RX MEDLINE=86108304; PubMed=3943519;
 RA Hoejrup P., Andersen S.O., Roepstorff P.;
 RT "Isolation, characterization, and N-terminal sequence studies of
 cuticular proteins from the migratory locust, Locusta migratoria.";
 RL Eur. J. Biochem. 154:153-159(1986).
 CC -1- FUNCTION: COMPONENT OF THE CUTICLE OF MIGRATORY LOCUST WHICH
 CONTAINS MORE THAN 100 DIFFERENT STRUCTURAL PROTEINS.
 CC -1- DOMAIN: THE TETRAPEPTIDE (A-A-P-[AV]) REPEATS FOUND THROUGHOUT THE
 PROTEIN ARE ALSO PRESENT IN MANY PROTEINS CONSTITUTING THE
 PROTECTIVE ENVELOPE OF OTHER SPECIES.
 CC -1- SIMILARITY: CONTAINS 1 CUTICLE CONSENSUS DOMAIN.

DR PIR; B24802; B24802.
 DR PIR; S05638; S05638.
 DR InterPro; IPR000618; Insect_cuticle.
 DR Pfam; PF00379; Insect_cuticle; 1.
 DR PRINTS; PR00947; CUTICLE.
 DR PROSITE; PS00233; CUTICLE; 1.
 KM Structural protein; Cuticle; Repeat.
 FT REPEAT 16 19 1.
 FT REPEAT 22 22 2.
 FT REPEAT 28 31 3.
 FT REPEAT 37 40 4.
 FT REPEAT 44 47 5.
 SO SEQUENCE 148 AA; 15224 MW; C6EADC27C593ACE6 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 148;
 Best Local Similarity 80.0%; Pred. No. 37;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
 |||||
 DB 64 NYGVH 68

RESULT 14

3MGL_RHIME STANDARD; PRT; 185 AA.
 AC Q92T11;
 DT 01-MAR-2002 (Rel. 41, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Putative 3-methyladenine DNA glycosylase (EC 3.2.2.-).
 GN R81416 OR S8B20709.
 OS Rhizidium melioli (Sinorhizidium melioli).
 OG Plasmid psymb (megaplasmid 2).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 NC NCBI_TaxID=382;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396508; PubMed=11481431;
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
 RA Vorhiesler F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
 RA Golding B., Puehler A.;
 RT "The complete sequence of the 1,683-kb psymb megaplasmid from the N2-
 fixing endosymbiont Sinorhizidium melioli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 CC -1- SIMILARITY: BELONGS TO THE DNA GLYCOSYLASE MFG FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AL0603647; CAC49816.1; -;
 DR Hypothetical protein; DNA repair; Hydrolase; Plasmid;
 KW Complete proteome.
 SO SEQUENCE 185 AA; 19974 MW; 7DE6283605F4A11 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 185;
 Best Local Similarity 80.0%; Pred. No. 47;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
 |||||
 DB 74 SYGLH 78

RESULT 15

GRP3_DAUCA STANDARD; PRT; 195 AA.
 AC P37705;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Glycine rich protein A3.
 OS Daucus carota (Carrot).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
 NC NCBI_TaxID=4039;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Schrader S., Kaldenhoff R., Richter G.;
 RL Submitted (MAR-1993) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC

DR EMBL; X72383; CA51076.1; -;
 DR PIR; S32123; S32123.
 SO SEQUENCE 195 AA; 19099 MW; 886078886107DAA5 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 195;
 Best Local Similarity 80.0%; Pred. No. 49;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
 :1111
 Db 115 AYGVH 119

Search completed: August 20, 2002, 13:17:51
 Job time: 260 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:13:26 ; Search time 73.7 Seconds
(Without alignments)
11.736 Million cell updates/sec

Title: US-09-824-286-3_COPY_28_32
Perfect score: 29
Sequence: 1 SYGVH 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_oranelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	100.0	86	9 Q9ZX39	Q9ZX39 mycobacteri
2	29	100.0	188	11 Q9DB11	Q9DB11 mus musculu
3	29	100.0	221	12 Q91416	Q91416 sulfolobus
4	29	100.0	240	16 Q99U88	Q99U88 staphylococ
5	29	100.0	285	2 Q9X5X8	Q9X5X8 pseudomonas
6	29	100.0	285	16 Q51514	Q51514 pseudomonas
7	29	100.0	288	16 Q99U13	Q99U13 staphylococ
8	29	100.0	367	2 Q87084	Q87084 helicobacte
9	29	100.0	392	5 Q9BL53	Q9BL53 leishmania
10	29	100.0	397	17 Q980C6	Q980C6 sulfolobus
11	29	100.0	461	2 Q9ANAS	Q9ANAS bradyrhizob
12	29	100.0	490	10 Q9FSS8	Q9FSS8 oryza sativ
13	29	100.0	631	5 Q97340	Q97340 photonis va
14	29	100.0	657	10 Q9STF3	Q9STF3 arabidopsis
15	29	100.0	703	10 Q9ST08	Q9ST08 brassica ca
16	29	100.0	806	16 Q9ZM29	Q9ZM29 helicobacte

17	29	100.0	821	5 Q9VL46	Q9VL46 drosophila
18	29	100.0	831	16 Q24906	Q24906 helicobacte
19	29	100.0	871	17 Q9HSM2	Q9HSM2 halobacteri
20	29	100.0	1124	5 Q9VYN6	Q9VYN6 drosophila
21	29	100.0	1274	3 Q9UT05	Q9UT05 schizosach
22	29	100.0	4881	2 Q9S0R3	Q9S0R3 streptomyce
23	28	96.6	82	7 Q9XR7	Q9XR7 rattus fusc
24	28	96.6	126	2 Q48988	Q48988 mycoplasma
25	28	96.6	133	12 Q41853	Q41853 duvenhage v
26	28	96.6	139	16 Q9KRF7	Q9KRF7 bacillus ha
27	28	96.6	176	16 Q98PA3	Q98PA3 rhizobium 1
28	28	96.6	200	16 Q98IJ3	Q98IJ3 rhizobium 1
29	28	96.6	289	16 Q9PHY0	Q9PHY0 campylobact
30	28	96.6	324	16 Q53481	Q53481 mycobacteri
31	28	96.6	434	2 Q9Z6C0	Q9Z6C0 enterobacte
32	28	96.6	450	12 Q82994	Q82994 lagos bat v
33	28	96.6	451	12 Q66453	Q66453 duvenhage v
34	28	96.6	600	10 Q41759	Q41759 zea mays (m
35	28	96.6	672	10 Q9LPA0	Q9LPA0 arabidopsis
36	28	96.6	804	10 Q94HX0	Q94HX0 oryza sativ
37	28	96.6	818	3 Q43059	Q43059 schizosach
38	28	96.6	841	2 Q93R04	Q93R04 lactococcus
39	28	96.6	860	16 Q92BV3	Q92BV3 listeria in
40	28	96.6	861	5 Q9VKT8	Q9VKT8 drosophila
41	28	96.6	1268	10 Q9SDC5	Q9SDC5 oryza sativ
42	28	96.6	1408	10 Q9FMU0	Q9FMU0 oryza sativ
43	28	96.6	1470	10 Q9AUP7	Q9AUP7 oryza sativ
44	28	96.6	1572	5 Q44938	Q44938 haemonchus
45	26	89.7	33	11 Q9QZ32	Q9QZ32 mus musculu

ALIGNMENTS

RESULT 1
ID Q9ZX39 PRELIMINARY; PRT; 86 AA.
AC Q9ZX39;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE GP39.
OS Mycobacteriophage TM4.
OC Viruses.
OX NCBI_TaxID=88870;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20110038; PubMed=10645443;
RA Ford M.E., Stenstrom C., Hendrix R.W., Hatfull G.F.;
RT "Mycobacteriophage TM4: Genome structure and gene expression.";
RL Tuber. Lung Dis. 79:63-73(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Ford M.E., Stenstrom C., Hendrix R.W., Hatfull G.F.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF068845; AADI7606.1;
SQ SEQUENCE 86 AA; 9303 MW; 77382ADCC082323 CRC64;

Query Match 100.0%; Score 29; DB 9; Length 86;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYGVH 5
| | | | |
Db 76 SYGVH 80

RESULT 2
ID Q9DB11 PRELIMINARY; PRT; 188 AA.
AC Q9DB11;
DT 01-JUN-2001 (TReMBLrel. 17, Created)

```
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE 2010001M09RIK PROTEIN.
GN 2010001M09RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Flieschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Botfeill D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilmink L.,
RA Wyshaw-Borls A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
DR EMBL; AK008016; BAB25410.1; -
DR MGD; MGI:1917066; 2010001M09RIK.
DR InterPro; IPR000886; ER_target.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
SQ SEQUENCE 188 AA; 20559 MW; C66F7BC3996A8278 CRC64;

Query Match
Best Local Similarity 100.0%; Score 29; DB 11; Length 188;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYGVH 5
Db 100 SYGVH 104

RESULT 3
091416 PRELIMINARY; PRT; 221 AA.
AC 091416;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 25.5 KDA PROTEIN.
GN SIFV0046.
OS Sulfolobus islandicus filamentous virus.
OC Viruses; dsDNA viruses, no RNA stage; Lipothirixviridae;
OC Lipothirixvirus.
OX NCBI_TaxID=176106;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20130740; PubMed=10662621;
RA Arnold H.P., Zillig W., Ziese U., Holz I., Crosby M., Uterback T.,
RA Weidmann J.F., Umayam L.A., Teffera K., Kristjanson J.K., Klenk H.P.,
RA Nelson K.E., Fraser C.M.;
RT "A novel lipothirixvirus, SIFV, of the extremely thermophilic
RT crenarchaeon Sulfolobus."
RL Virology 267:252-266(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Arnold H.P., Zillig W., Ziese U., Holz I., Crosby M., Uterback T.,
```

```
RA Weidmann J.F., Umayam L.A., Teffera K., Kristjanson J.K., Klenk H.P.,
RA Nelson K.E., Fraser C.M.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF440571; AAL27755.1; -.
KW Hypothetical protein.
SQ SEQUENCE 221 AA; 25486 MW; C278CC99CDF11D15 CRC64;

Query Match
Best Local Similarity 100.0%; Score 29; DB 12; Length 221;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYGVH 5
Db 69 SYGVH 73

RESULT 4
099088 PRELIMINARY; PRT; 240 AA.
ID 099088
AC 099088;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE DIHYDRODIPICOLINATE REDUCTASE.
GN DABP OR SA1228 OR SAV1396.
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879, 158878;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hoshiyama A.,
RA Mizutani Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiratsuka K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003133; BAB42488.1; -.
DR EMBL; AP003362; BAB57558.1; -.
DR HSSP; P04036; IDRM.
DR InterPro; IPR000846; DABP.
DR Pfam; PF01113; DABP; 1.
DR Prodom; PD004105; DABP; 1.
KW Complete proteome.
SQ SEQUENCE 240 AA; 26726 MW; F910FBD3DBF4B8D CRC64;

Query Match
Best Local Similarity 100.0%; Score 29; DB 16; Length 240;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYGVH 5
Db 107 SYGVH 111

RESULT 5
09X5X8 PRELIMINARY; PRT; 285 AA.
ID 09X5X8
AC 09X5X8;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PHA DEPOLYMERASE.
GN PHA2.
```


OS Pseudomonas resinovorans.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=53412;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL B-2649;
RA Solaiman D.K.Y.;
RT "PCR cloning of Pseudomonas resinovorans polyhydroxyalkanoate biosynthesis genes and expression in Escherichia coli.";
RL Biotechnol. Lett. 22:789-794(2000).
DR EMBL: AF129396; AAD26366.1; -;
DR InterPro: IPR000073; Abhydrolase.
DR InterPro: IPR003088; AB_hydrolase.
DR InterPro: IPR000379; Est_lip_thioest_actsite.
DR Pfam: PF00561; abhydrolase; 1.
DR PRINTS: PR00111; ABHYDROLASE.
DR PROSITE: PS00120; LIPASE_SER; UNKNOWN_1.
SQ SEQUENCE 285 AA; 31463 MW; F23856753C801B0E CRC64;

Query Match 100.0%; Score 29; DB 2; Length 285;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 154 SYGVH 158

RESULT 6
ID 051514 PRELIMINARY; PRT; 285 AA.
AC 051514;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PPA-DEPOLYMERASE (POLY(3-HYDROXYALKANOIC ACID) DEPOLYMERASE).
GN PHAD OR PA5057.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RA Steinbuechel A.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=93011120; PubMed=1396693;
RA Timm A.; Steinbuechel A.;
RT "Cloning and molecular analysis of the poly(3-hydroxyalkanoic acid) gene locus of Pseudomonas aeruginosa PA01.";
RL Eur. J. Biochem. 209:15-30(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K.; Pham X.-Q.T.; Erwin A.L.; Mizoguchi S.D.; Warren P.; Hickey M.J.; Brinkman F.S.L.; Huftagle W.O.; Kowalik D.J.; Lagrou M.; Garber R.L.; Goltzy L.; Tolentino E.; Westbrock-Wadman S.; Yuan Y.; Brody L.L.; Coulter S.N.; Folger K.R.; Kas A.; Larbig K.; Lim R.M.; Smith K.A.; Spencer D.H.; Mong G.K.-S.; Wu Z.; Paulsen I.T.; Reizer J.; Sailer M.H.; Hancock R.E.W.; Lory S.; Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR EMBL: X66592; CAA47152.1; -;
DR EMBL: AE004919; AAG08442.1; -;

DR InterPro: IPR000073; Abhydrolase.
DR InterPro: IPR003088; AB_hydrolase.
DR InterPro: IPR000379; Est_lip_thioest_actsite.
DR Pfam: PF00561; abhydrolase; 1.
DR PRINTS: PR00111; ABHYDROLASE.
KW Complete proteome.
SQ SEQUENCE 285 AA; 31470 MW; C384C7D6435D5C6C CRC64;

Query Match 100.0%; Score 29; DB 16; Length 285;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 154 SYGVH 158

RESULT 7
ID 099013 PRELIMINARY; PRT; 288 AA.
AC 099013;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SA1132 PROTEIN (HYPOTHETICAL PROTEIN SAV1290).
GN SA1132 OR SAV1290.
OS Staphylococcus aureus (strain N315), and
OC Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879, 158878;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I., Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T., Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C., Sekimizu K., Hirakawa H., Kubara S., Goto S., Yabuzaki J., Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003133; BAB42385.1; -;
DR EMBL: AP003361; BAB57452.1; -;
DR InterPro: IPR000399; TPP_enzyme.
DR Pfam: PF02775; TPP_enzymes_C; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 288 AA; 31331 MW; 3A2AD0882C7FBCD4 CRC64;

Query Match 100.0%; Score 29; DB 16; Length 288;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
|||||
Db 60 SYGVH 64

RESULT 8
ID 087084 PRELIMINARY; PRT; 367 AA.
AC 087084;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CELL DIVISION-RELATED GENE.
GN CDRA.
OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HPK5;
RX MEDLINE=98422487; PubMed=9748467;
RA Takeuchi H., Shirai M., Akada J.K., Tsuda M., Nakazawa T.;
RT "Nucleotide sequence and characterization of cdra, a cell division-
related gene of Helicobacter pylori.";
RL J. Bacteriol. 180:5263-5268(1998).
DR EMBL: AB003309; BAA33499.1; -;
DR InterPro: IPR002543; FtsK_SpoIIIE.
DR Pfam: PF01580; FtsK_SpoIIIE; 1.
SQ SEQUENCE 367 AA; 42500 MW; C6D43C01D6DB1434 CRC64;

Query Match 100.0%; Score 29; DB 2; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYGVH 5
DB 321 SYGVH 325

RESULT 9
Q9BLS3 PRELIMINARY; PRT; 392 AA.
ID Q9BLS3
AC Q9BLS3;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE HYPOTHETICAL 41.1 KDA PROTEIN.
GN L6071.08.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Zimmermann W., Ivens A.C., Quail M., Rajandream M.A., Barrell B.G.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RT Genome Res. 8:135-145(1998).
DR EMBL: AL583933; CAC32267.1; -;
KW Hypothetical protein.
SQ SEQUENCE 392 AA; 41087 MW; 4A7DC7FC291785E3 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYGVH 5
DB 231 SYGVH 235

RESULT 10
Q980C6 PRELIMINARY; PRT; 397 AA.
ID Q980C6
AC Q980C6;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
DE HYPOTHETICAL PROTEIN SSC00390.

GN SSC00390.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Conlathierl F., Zivanovic Y., Allard G.,
RA Aweez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL: AE006672; AKA40717.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 397 AA; 41545 MW; 01C2618BA0D7C2DA CRC64;

Query Match 100.0%; Score 29; DB 17; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYGVH 5
DB 237 SYGVH 241

RESULT 11
Q9ANAS PRELIMINARY; PRT; 461 AA.
ID Q9ANAS
AC Q9ANAS;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
DE ID397.
GN ID397.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=110SPC4;
RX MEDLINE=21101824; PubMed=1157954;
RA Gottfert M., Rothlisberger S., Kundig C., Beck C., Marty R.,
RA Hennecke H.;
RT "Potential symbiosis-specific genes uncovered by sequencing a 410-kb
RT DNA region of the Bradyrhizobium japonicum chromosome.";
RT J. Bacteriol. 183:1405-1412(2001).
DR EMBL: AF322012; AAG60868.1; -;
DR InterPro: IPR000567; SBP_bac1.
DR Pfam: PF01547; SBP_bacterial_1; 1.
SQ SEQUENCE 461 AA; 50616 MW; 5043D981CC62A771 CRC64;

Query Match 100.0%; Score 29; DB 2; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYGVH 5
DB 240 SYGVH 244

RESULT 12
Q9FSS8 PRELIMINARY; PRT; 490 AA.
ID Q9FSS8
AC Q9FSS8;
DT 01-MAR-2001 (TRENBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE PTNATIVE WALL-ASSOCIATED KINASE 2.
 GN H0212B02.2.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Erihartoideae; Oryzaceae; Oryza.
 NC NCB1_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Han B., Zhou B., Feng Q., Huang Y.C., Chen Z.H., Li Y., Zhu J.J.,
 RA Tang Y.S., Zhao Q., Liu Y.L., Mu J., Yu Z., Fan D.L., Chen L.,
 RA Wang Q.J., Zhang L., Lu Y.Q., Yu S.L., Zhu J., Liu X.H., Hu X.,
 RA Lei H.Y., Zhang Y.J., Wang R., Li C., Lu Y., Chen X.C., Zhang Y.,
 RA Hu H., Jia P.X., Li T., Qian Y.M., Ying K., Hong G.F.,
 RA "Oryza sativa indica (Guangluai4) genomic DNA, chromosome4, BAC
 RT clone:H0212B02."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL442007; CAC09346.1; -
 KW Kinase.
 SO SEQUENCE 490 AA; 52039 MW; 2F1301B08BA26EFE CRC64;

Query Match 100.0%; Score 29; DB 10; Length 490;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SYGVH 5
 |||||
 DB 186 SYGVH 190

RESULT 13
 097340
 ID 097340 PRELIMINARY; PRT; 631 AA.
 AC 097340;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CYTOPLASMIC INTERMEDIATE FILAMENT PROTEIN.
 OS Photons vanconverensis.
 OC Eukaryota; Metazoa; Brachlopoda; Photoniiformes; Photons.
 NC NCB1_TaxID=34521;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99065768; Pubmed=9847417;
 RA Eder A., Riemer D., Bovenshulte M., Weber K.;
 RT "Molecular Phylogeny of metazoan intermediate filament proteins."
 RL J. Mol. Evol. 47:751-762(1998).
 DR EMBL; AJ004936; CAB38181.1; -
 DR InterPro: IPR001322; IF_tail.
 DR InterPro: IPR001322; IF_tail.
 DR Pfam: PF00038; filament; 1.
 DR Pfam: PF00932; IF_tail; 1.
 DR PROSITE: PS00226; IF; UNKNOWN_1.
 SO SEQUENCE 631 AA; 71028 MW; 64856BEF7D82379 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 631;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
 |||||
 DB 71 SYGVH 75

RESULT 14
 09STF3
 ID 09STF3 PRELIMINARY; PRT; 657 AA.
 AC 09STF3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHEICAL 74.3 KDA PROTEIN.
 GN T6H20.180.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCB1_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Choise N., Robert C., Brothier P., Wincker P., Catolico L.,
 RA Arigenave F., Saurin W., Weissenbach J., Mewes H.W., Mayer K.F.X.,
 RA Lemcke K., Schueller C., Queller F., Salanoubat M.;
 RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL096859; CAB51186.1; -
 DR InterPro: IPR002885; PPR.
 DR Pfam: PF01535; PPR; 6.
 DR Hypothetical protein.
 KW
 SO SEQUENCE 657 AA; 74340 MW; 8257019E45CCDFC8 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 657;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 SYGVH 5
 |||||
 DB 360 SYGVH 364

RESULT 15
 09ST08
 ID 09ST08 PRELIMINARY; PRT; 703 AA.
 AC 09ST08;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE TRANSPOSON-LIKE ORF.
 GN TRANSPOSON-LIKE ORF.
 OS Brassica campestris (field mustard).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 NC NCB1_TaxID=3711;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99403015; Pubmed=10471721;
 RA Suzuki G., Kai N., Hirose T., Fukui K., Nishio T., Takayama S.,
 RA Isogai A., Watanabe M., Hinata K.;
 RT "Genomic organization of the S locus: identification and
 RT characterization of genes in SlG/SRK region of S9 haplotype of
 RT Brassica campestris (syn. rapa).";
 RL Genetics 153:391-400(1999).
 DR EMBL; AB022082; BAB85462.1; -
 DR InterPro: IPR004242; Transposase_21.
 DR Pfam: PF02992; Transposase_21; 1.
 SO SEQUENCE 703 AA; 81881 MW; AC46541BEC83C79 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 703;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5
 |||||
 DB 546 SYGVH 550

Aug 20 13:05:57 2002

us-09-824-286-3_copy_28_32.rspt

Page 6

Search completed: August 20, 2002, 13:17:23
Job time: 237 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:15:09 ; Search time 94.22 Seconds
(without alignments)
17.683 Million cell updates/sec

Title: US-09-824-286-3_COPY_47_61
Perfect score: 80
Sequence: 1 VIMAGSGTMYNSALM 15

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	100.0	16	AAAB1965	Ganglioside GD2 sp
2	80	100.0	16	AAAG3989	Monoclonal antibody
3	80	100.0	112	AAAW1648	Monoclonal antibody
4	80	100.0	113	AAAR2168	Murine VH group 2
5	80	100.0	116	AAAR7322	Murine VH group 2
6	80	100.0	116	AAW14450	Monoclonal antibody
7	80	100.0	116	AAAR9877	Monoclonal antibody
8	80	100.0	118	AAAB1970	Ganglioside GD2 sp
9	80	100.0	118	AAAB1973	Ganglioside GD2 sp
10	80	100.0	120	AAAG3986	Amino acid sequence
11	80	100.0	137	AAAB1975	Ganglioside GD2 sp

12	80	100.0	222	14	AAAR32843	VH NQ2/12.4-Vk NQ1
13	80	100.0	235	14	AAAR32840	VH NQ2/12.4-Vk NQ1
14	80	100.0	269	13	AAAR32569	Fusion protein enc
15	80	100.0	581	22	AAAB1972	Ganglioside GD2 sp
16	77	96.2	142	22	AAAG6520	Mouse antibody 26
17	77	96.2	142	22	AAAG6523	Humanised anti-CTL
18	77	96.2	239	17	AAAR3679	Single chain polyp
19	77	96.2	239	17	AAW02191	18-2-3/TRY202' sin
20	77	96.2	239	17	AAAR9649	Single chain bindi
21	77	96.2	241	11	AAAR06482	18-2-3/TRY202' . A
22	77	96.2	242	11	AAAR06483	18-2-3/TRY59. A
23	77	96.2	242	14	AAAR3680	Single chain polyp
24	77	96.2	242	17	AAW02192	18-2-3/TRY59 singl
25	77	96.2	242	17	AAAR9650	Single chain bindi
26	76	95.0	272	18	AAW16688	Murine anti-human
27	74	92.5	140	18	AAW2538	Murine anti-human
28	71	88.8	29	18	AAW13989	Single chain antib
29	71	88.8	29	18	AAW13988	Single chain antib
30	71	88.8	29	18	AAW13987	Single chain antib
31	71	88.8	29	18	AAW13986	Single chain antib
32	71	88.8	476	22	AAAB49243	Chimeric 4H6 anti-
33	70	87.5	119	21	AAAR0818	260F9 hybridoma VL
34	69.5	86.9	115	18	AAW04595	Anti-DNA antibody
35	69.5	86.9	123	18	AAW07438	Anti-DNA antibody
36	66	82.5	231	18	AAW27090	Mouse monoclonal a
37	65	81.2	16	17	AAAR8481	Anti-IL-5 Mab heav
38	65	81.2	111	17	AAAR8494	Mab 2E3 heavy cha
39	65	81.2	111	17	AAAR8496	Mab 2E3 heavy cha
40	65	81.2	111	19	AAW42453	Mouse anti-human I
41	65	81.2	111	19	AAW42455	Mouse anti-human I
42	65	81.2	119	17	AAAR8478	Mab 2B6 heavy cha
43	65	81.2	119	17	AAAR8488	Humanised 2B6 anti
44	65	81.2	119	17	AAAR8492	Humanised 2B6 anti
45	65	81.2	119	19	AAW42465	Interleukin-5 huma

ALIGNMENTS

RESULT 1
ID AAB81965 standard; Peptide: 16 AA.
XX AAB81965;
XX 03-JUL-2001 (first entry)
XX Ganglioside GD2 specific antibody related peptide SEQ ID NO: 4.
DE Ganglioside; GD2; complementation determining region; CDR; antibody;
XX Ganglioside; GD2; complementation determining region; CDR; antibody;
KW mouse; cancer.
XX
XX OS Mus musculus.
XX WO200123573-A1.
XX 05-APR-2001.
XX 29-SEP-2000; 2000WO-JP06773.
XX 30-SEP-1999; 99JP-0278290.
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX Hanai N, Shtata K, Nakamura K, Niwa R;
XX WPI; 2001-266163/27.
XX Human type complementation-determining domain transplanted antibody and
XX derivatives against ganglioside GD2, useful in diagnosis and therapy of
XX e.g. tumors, has low antigenicity, little side effects but potent
XX activity in cancer

PS Claim 5; Page 99; 123pp; Japanese.
XX
CC The present invention describes an antibody, which can react specifically
CC with ganglioside GD2, and is transplanted with a human type
CC complementation-determining domain (CDR), or its fragments. The antibody
CC and its derivatives are useful in diagnosis and therapy of tumours,
CC particularly cancer diagnosis. The present sequence is a peptide
CC used in the exemplification of the invention.
XX
SQ Sequence 16 AA;

Query Match 100.0%; Score 80; DB 22; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 VIMAGSTNYNSALM 15
|||||
Db 1 vimagstnynsalm 15

RESULT 2
AAG63989
ID AAG63989 standard; peptide; 16 AA.

AC AAG63989;
XX
DT 26-NOV-2001 (first entry)

DE Complementarity determining region of heavy chain of antibody 2C4.

KW Monoclonal antibody 2C4; staloadesin factor-2; SAF-2; allergic rhinitis;
KW allergy; asthma; anemia; eczema; lymphoma; systemic mastocytosis;
KW leukemia; eosinophil.

OS Mus sp.

PN WO200166126-A1.

PD 13-SEP-2001.

PF 05-MAR-2001; 2001WO-US07193.

PR 07-MAR-2000; 2000US-0187595.

XX (SMIR) SMITHKLINE BEECHAM CORP.

PA (SMIR) SMITHKLINE BEECHAM PLC.

PA (UYJO) UNIV JOHNS HOPKINS.

XX
PI Abrahamson JA, Bochner B, Erickson-Miller CL, Kikly KK;
PI Schleimer R;

DR WPI; 2001-570749/64.

XX Novel monoclonal antibody specific for human staloadesin factor-2 for
PT diagnosis, prevention, treatment of allergy, asthma, eczema or diseases
PT such as lymphoma, leukemia or systemic mastocytosis, in a mammal -
XX
PS Claim 10; Page 33; 35pp; English.

XX AAG63988-90 represent the complementarity determining regions (CDRs)
CC of the heavy chain variable region of murine monoclonal antibody 2C4.

CC This antibody binds to human staloadesin factor-2 (SAF-2). The
CC antibody is useful for treating or preventing allergic rhinitis,
CC allergies, asthma, anemia, eczema or diseases such as lymphoma,
CC leukemia or systemic mastocytosis in a mammal. It is also useful for
CC detecting the presence of a cell, especially eosinophil in a sample,
CC by detecting binding of the antibody to SAF-2. The antibody can be
CC coupled to toxins, antiproliferative drugs or radionuclides to
CC kill cells in areas of excessive SAF-2 expression.

XX Sequence 16 AA;

Query Match 100.0%; Score 80; DB 22; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 VIMAGSTNYNSALM 15
|||||
Db 1 vimagstnynsalm 15

RESULT 3
AAW31648
ID AAW31648 standard; Protein; 112 AA.

AC AAW31648;

DT 21-MAY-1998 (first entry)

DE Monoclonal antibody CP.B8 heavy chain variable region.

XX Cytokine receptor; gamma common chain; gc chain; human;
KW blocking agent; monoclonal antibody; CP.B8; immunological disease;
KW myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
KW insulin-dependent diabetes; inflammatory bowel disease;
KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;
KW graft versus host disease; psoriasis; immunosuppressive; therapy;
KW complementarity determining region; CDR.

OS Mus musculus.

PN WO9743416-A1.

PD 20-NOV-1997.

PF 09-MAY-1997; 97WO-US07870.

PR 10-MAY-1996; 96US-0017466.

XX (BIOJ) BIOGEN INC.

XX Benjamin CD, Burkly LC, Hession C, Whitty A;

XX WPI; 1998-008885/01.

XX N-PSDB; AAT97441.

XX Blocking agents of the gamma common chain of cytokine receptors -
PT particularly monoclonal antibodies, used to induce T cell anergy for
PT treatment of immunological diseases
XX
PS Claim 23; Page 81-82; 111pp; English.

XX This polypeptide comprises the heavy chain variable region (VH) of
CC monoclonal antibody (Mab) C9.B8, which is produced by a hybridoma
CC deposited as ATCC 12107, and which is specific for the gamma
CC constant (gc) chain (see AAW31646) of human cytokine receptors. The
CC invention provides compositions and methods for inhibiting cytokine
CC signalling using gc chain blocking agents for the treatment of
CC immunological diseases such as myasthenia gravis, rheumatoid
CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,
CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,
CC allergy, asthma, parasitic infection, graft vs. host disease or
CC psoriasis. Preferred gc blocking agents include Mab CP.B8, its Fab
CC fragment and an antibody having a light chain variable region
CC CDR selected from those of CP.B8 VH or a heavy chain variable
CC region CDR selected from those of CP.B8 VL (see AAW31647).

```
XX SQ Sequence 112 AA;
Query Match 100.0%; Score 80; DB 19; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VIMAGSTNYNSALM 15
   |||||||
DB 47 viwagstnynsalm 61

RESULT 4
AAR21268
ID AAR21268 standard; Protein: 113 AA.
XX AC AAR21268;
XX DT 21-MAY-1992 (first entry)
XX DE Murine VH group 2 chain E specific for phox.
XX KW Fg; bacteriophage; gene III; filamentous; phagemid; capsid; coat;
XX KW pilus; g3p; binding; adsorption; gene VIII; diverse repertoire;
XX KW specific binding pairs; replicable genetic display package.
XX OS Synthetic.
XX FH Key
FH FT Binding-site Location/Qualifiers
FT FT /label= CDR1 31..35
FT FT Binding-site 50..65
FT FT /label= CDR2 98..102
FT FT Binding-site /label= CDR3
FT FT /note= "D/N-X-G-X-X motif "
FT FT
XX PN WO9201047-A.
XX PD 23-JAN-1992.
XX PF 10-JUL-1991; 91WO-GB01134.
XX PR 15-MAY-1991; 91GB-0010549.
XX PR 10-JUL-1990; 90GB-0015198.
XX PR 19-OCT-1990; 90GB-0022845.
XX PR 12-NOV-1990; 90GB-0024503.
XX PR 06-MAR-1991; 91GB-0004744.
XX PA (CAMP-) CAMBRIDGE ANTIBODY.
XX PA (MEDI-) MED RES COUNCIL.
XX PI McCafferty J, Pope AR, Johnson KS, Hoogenboom HRJ, Griffiths AD;
XX PI Jackson RH, Holliger KP, Marks JD, Clackson TP, Chiswell DJ;
XX PI Winter GP, Bonfert TP;
XX DR WPI: 1992-056862/07.
XX PT Producing members of specific binding pairs - by expression in
XX PT recombinant host cells with a secreting replicable genetic
XX PT display package.
XX PS Example 21: Fig 24; 209pp; English.
XX CC The VH sequence is one of eight (AAR21264-71) found to be expressed
XX CC from a single chain Fv library from an immunised mouse. The
XX CC library produces a diverse repertoire of antibody fragments specific
XX CC for 2-phenyl-5-oxazolone (phox). It was prep'd. using cDNA generated
XX CC from mRNA from mice immunised with phox coupled to chicken serum
XX CC albumin. The VH and VL kappa sequences were separately amplified by
XX CC PCR (see AAQ23474-84) and ligated into fdCAT2 (see AAQ23463) for
XX CC expression on the phage surface as fusions with gene III. The
```

```
CC resulting library of clones was diverse. Twenty three hapten binding
CC clones were sequenced revealing the eight different VH genes (A-H)
CC in a variety of pairings with seven different VK genes (a-g) (see
CC AAR21264-92). Nearly all the VH genes belonged to gp 1, with only
CC one, "E", being of gp 2 (VHox1). Of the twenty three clones
CC sequenced, only one was of type "E". Most of the clones were VK-d
CC combinations. The Kd of VH-B/VK-d for phox-GABA was 10 nM. Only two
CC other combinations (of eleven tested) were found to have higher
CC values. This suggests that phage bearing scFv fragments having
CC weak affinities can be selected with antigen, probably due to the
CC avidity of the multiple antibody heads on the phage.
CC See also AAR21260-307, 309-311; AAR22450, 565-581.
XX SQ Sequence 113 AA;
Query Match 100.0%; Score 80; DB 13; Length 113;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VIMAGSTNYNSALM 15
   .|||||
DB 50 viwagstnynsalm 64

RESULT 5
AAR07322
ID AAR07322 standard; protein: 116 AA.
XX AC AAR07322;
XX DT 22-JAN-1991 (first entry)
XX DE VH domain of antibody D against tumour-associated antigens.
XX KW Tumour-associated antigen; murine monoclonal antibody D;
XX KW vibrio cholerae neuraminidase-sensitive epitope; glioma;
XX KW meningioma; neurilemmoma; neuroblastoma; ganglioblastoma;
XX KW ganglioneuroma; diagnosis.
XX OS Mus musculus.
XX PN EP388914-A.
XX PD 26-SEP-1990.
XX PF 21-MAR-1990; 90EP-0105322.
XX PR 24-MAR-1989; 89DE-3909799.
XX PA (BEHM ) BEHRINGWERKE AG.
XX PI Bosslet K, Seemann G, Sedlacek HH;
XX PI WPI: 1990-291873/39.
XX DR N-PSDB; AAQ06229.
XX PT Monoclonal antibodies to tumour associated antigens - used for
XX PT diagnosis of malignant tumours etc.
XX PS Disclosure; Page 14; 18pp; German.
XX CC Antibody D is produced as described in EP-141079 and binds to
XX CC a vibrio cholerae neuraminidase-sensitive epitope on ganglioside Gp2
XX CC which occurs in gliomas, meningiomas, neurilemmomas,
XX CC neuroblastomas, ganglioblastomas and ganglioneuromas.
XX CC They are useful in tumour diagnosis and therapy.
XX CC See also AAQ06230 for VK of Mab D, AAQ07312-15 for Mab A and B and
XX CC AAQ06227-28 for Mab C.
XX SQ Sequence 116 AA;
```

Query Match 100.0%; Score 80; DB 11; Length 116;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15
 |||||
 DB 45 vlmagstnynsalm 59

RESULT 6
 AAM14490
 ID AAM14490 standard; Protein; 116 AA.
 XX
 AC AAM14490;
 XX
 DT 28-JAN-1997 (first entry)
 XX
 DE Monoclonal antibody D VH.
 XX
 KW heavy; light chain; monoclonal antibody; antigen 3; marker; melanoma;
 KW permanent human tumour cell line; tumour-associated antigen; epitope;
 KW gastrointestinal tumour; pancreatic carcinoma; diagnostic; therapeutic;
 KW antigen 11; Vibrio cholera; neuraminidase-resistant; ganglioside GD2.
 XX
 OS Synthetic.
 XX
 PN EP727436-A1.
 XX
 PD 21-AUG-1996.
 XX
 PE 21-MAR-1990; 90EP-0105322.
 XX
 PR 24-MAR-1989; 89DE-3909799.
 XX
 PA (BEHW) BEHRINGWERKE AG.
 XX
 PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;
 DR WPI: 1996-372836/38.
 DR N-PSDB; AAT63507.
 XX
 PT Monoclonal antibody to tumour-associated antigen - useful as
 PT gastrointestinal tumour marker
 PS
 PS Disclosure: Page 14; 19pp; German.
 XX
 CC AAM14490-91 are the heavy and light chains (respectively) of monoclonal
 CC antibody (Mab) D. Mab D recognises Vibrio cholera
 CC neuraminidase-resistant epitope of ganglioside GD2, from a human melanoma
 CC cell line. Mabs A, B and C (see AAM14484-89) are mentioned in the
 CC specification, but are not part of the claims. Mabs A and B recognise
 CC antigens 3 and 11 resp., of a permanent human tumour cell line. Mab C
 CC also recognises an epitope of a tumour-associated antigen. These antigens
 CC occur at high concns. in the serum of patients with gastrointestinal
 CC tumours, e.g. pancreatic carcinoma, and are thus useful as tumour markers
 CC for diagnostic or therapeutic purposes.
 XX
 SQ Sequence 116 AA;

Query Match 100.0%; Score 80; DB 17; Length 116;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15
 |||||
 DB 45 vlmagstnynsalm 59

RESULT 7
 AAR9877
 ID AAR9877 standard; Protein; 116 AA.
 XX

AC AAR9877;
 XX
 DT 28-JAN-1997 (first entry)
 XX
 DE Monoclonal antibody D VH.
 XX
 KW Monoclonal antibody; Mab; epitope; tumour-associated antigen;
 KW marker; antigen.
 XX
 OS Synthetic.
 XX
 PN EP727435-A1.
 XX
 PD 21-AUG-1996.
 XX
 PE 21-MAR-1990; 90EP-0105322.
 XX
 PR 24-MAR-1989; 89DE-3909799.
 XX
 PA (BEHW) BEHRINGWERKE AG.
 XX
 PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;
 DR WPI: 1996-372835/38.
 DR N-PSDB; AAT3665.
 XX
 PT Monoclonal antibody to tumour-associated antigen - useful as
 PT gastrointestinal tumour marker
 PS
 PS Disclosure: Page 14; 19pp; German.
 XX
 CC Mab C (AAT36659-T36660) is a monoclonal antibody that recognises an
 CC epitope of a tumour-associated antigen occurring at high concn. in
 CC the serum of patients with gastrointestinal tumours, e.g. pancreatic
 CC carcinoma, and is thus useful as a tumour marker for diagnostic or
 CC therapeutic purposes.
 CC
 CC Mabs A, B and D are mentioned in the specification, but are not
 CC part of the claims.
 CC Mab A (AAT36661-T36662) recognises antigen 3 of permanent human
 CC tumour cell line.
 CC Mab B (AAT36663-T36664) recognises antigen 11 of permanent human
 CC tumour cell line.
 CC Mab D (AAT36665-T36666) recognises a Vibrio cholera neuraminidase-
 CC resistant epitope of ganglioside GD2, from a human melanoma cell
 CC line.
 XX
 SQ Sequence 116 AA;

Query Match 100.0%; Score 80; DB 17; Length 116;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15
 |||||
 DB 45 vlmagstnynsalm 59

RESULT 8
 AAB81970
 ID AAB81970 standard; Protein; 118 AA.
 XX
 AC AAB81970;
 XX
 DT 03-JUL-2001 (first entry)
 XX
 DE Ganglioside GD2 specific antibody related protein SEQ ID NO: 13.
 XX
 KW Ganglioside; GD2; complementation determining region; CDR; antibody;
 KW mouse; cancer.
 XX
 OS Synthetic.
 XX

PN WO200123573-A1.
 XX
 PD 05-APR-2001.
 XX
 PE 29-SEP-2000; 2000WO-JP06773.
 XX
 PR 30-SEP-1999; 99JP-0278290.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Hanai N, Shitara K, Nakamura K, Niwa R;
 XX
 DR WPI; 2001-266163/27.
 XX
 PT Human type complementation-determining domain transplanted antibody and
 PT derivatives against ganglioside GD2, useful in diagnosis and therapy of
 PT e.g. tumours, has low antigenicity, little side effects but potent
 PT activity in cancer -
 XX
 PS Claim 10; Page 102-103; 123pp; Japanese.
 XX
 CC The present invention describes an antibody, which can react specifically
 CC with ganglioside GD2, and is transplanted with a human type
 CC complementation-determining domain (CDR), or its fragments. The antibody
 CC and its derivatives are useful in diagnosis and therapy of tumours,
 CC particularly cancer diagnosis. The present sequence is a protein
 CC used in the exemplification of the invention.
 CC
 SQ Sequence 118 AA;

Query Match 100.0%; Score 80; DB 22; Length 118;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
 |||||
 Db 50 viwagstnynsalm 64

RESULT 9
 AAB81973
 ID AAB81973 standard; Protein; 118 AA.
 XX
 AC AAB81973;
 XX
 DT 03-JUL-2001 (first entry)
 XX
 DE Ganglioside GD2 specific antibody related protein SEQ ID NO: 32.
 XX
 KM Ganglioside; GD2; complementation determining region; CDR; antibody;
 KM mouse; cancer.
 XX
 OS Mus musculus.
 XX
 PN WO200123573-A1.
 XX
 PD 05-APR-2001.
 XX
 PE 29-SEP-2000; 2000WO-JP06773.
 XX
 PR 30-SEP-1999; 99JP-0278290.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Hanai N, Shitara K, Nakamura K, Niwa R;
 XX
 DR WPI; 2001-266163/27.
 XX
 PT Human type complementation-determining domain transplanted antibody and
 PT derivatives against ganglioside GD2, useful in diagnosis and therapy of
 PT e.g. tumours, has low antigenicity, little side effects but potent
 PT activity in cancer -
 XX
 PS Sequence 120 AA;

XX Claim 26; Page 115-116; 123pp; Japanese.
 PS
 XX
 CC The present invention describes an antibody, which can react specifically
 CC with ganglioside GD2, and is transplanted with a human type
 CC complementation-determining domain (CDR), or its fragments. The antibody
 CC and its derivatives are useful in diagnosis and therapy of tumours,
 CC particularly cancer diagnosis. The present sequence is a protein
 CC used in the exemplification of the invention.
 CC
 SQ Sequence 118 AA;

Query Match 100.0%; Score 80; DB 22; Length 118;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
 |||||
 Db 50 viwagstnynsalm 64

RESULT 10
 AAG63986
 ID AAG63986 standard; Protein; 120 AA.
 XX
 AC AAG63986;
 XX
 DT 26-NOV-2001 (first entry)
 XX
 DE Amino acid sequence of heavy chain variable region of antibody 2c4.
 XX
 KM Monoclonal antibody 2c4; sialoadhesin factor-2; SAF-2; allergic rhinitis;
 KM allergy; asthma; anemia; eczema; lymphoma; systemic mastocytosis;
 KM leukemia; eosinophil.
 XX
 OS Mus sp.
 XX
 PN WO200166126-A1.
 XX
 PD 13-SEP-2001.
 XX
 PE 05-MAR-2001; 2001WO-US07193.
 XX
 PR 07-MAR-2000; 2000US-0187595.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PA (UYJO) UNIV JOHNS HOPKINS.
 XX
 PI Abrahamson JA, Bochner B, Erickson-Miller CL, Kikly KK;
 PI Schleimer R;
 XX
 DR WPI; 2001-570749/64.
 DR N-PSDB; AAH78183.
 XX
 PT Novel monoclonal antibody specific for human sialoadhesin factor-2 for
 PT diagnosis, prevention, treatment of allergy, asthma, eczema or diseases
 PT such as lymphoma, leukemia or systemic mastocytosis, in a mammal -
 XX
 PS Claim 17; Fig 1; 35pp; English.
 XX
 CC The present sequence represents the heavy chain variable region of murine
 CC monoclonal antibody 2c4. This antibody binds to human sialoadhesin
 CC factor-2 (SAF-2). The antibody is useful for treating or preventing
 CC allergic rhinitis, allergies, asthma, anemia, eczema or diseases such
 CC as lymphoma, leukemia or systemic mastocytosis in a mammal. It is also
 CC useful for detecting the presence of a cell, especially eosinophil in
 CC a sample, by detecting binding of the antibody to SAF-2. The antibody
 CC can be coupled to toxins, antiproliferative drugs or radionuclides to
 CC kill cells in areas of excessive SAF-2 expression.
 CC
 SQ Sequence 120 AA;

Query Match 100.0%; Score 80; DB 22; Length 120;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGGSTNYSALM 15
 |||||||
 DB 50 vimaggstnysalm 64

RESULT 11

AAB81975
 ID AAB81975 standard; Protein; 137 AA.

XX AAB81975;

XX 03-JUL-2001 (first entry)

XX Ganglioside GD2 specific antibody related protein #1.

XX Ganglioside; GD2; complementation determining region; CDR; antibody;

KW mouse; cancer.

XX Mus musculus.

PN WO200123573-A1.

PD 05-APR-2001.

PE 29-SEP-2000; 2000WO-JP06773.

PR 30-SEP-1999; 99JP-0278290.

PA (KYOW) KYOWA HAKKO KOGYO KK.

PI Hanai N, Shitara K, Nakamura K, Niwa R;

DR WPI; 2001-266163/27.

DR N-PSDB; AAF86854.

PT Human type complementation-determining domain transplanted antibody and
 PT derivatives against ganglioside GD2, useful in diagnosis and therapy of
 PT e.g. tumours, has low antigenicity, little side effects but potent
 PT activity in cancer -

XX Example 2; Page 96-97; 123pp; Japanese.

XX The present invention describes an antibody, which can react specifically
 CC with ganglioside GD2, and is transplanted with a human type
 CC complementation-determining domain (CDR), or its fragments. The antibody
 CC and its derivatives are useful in diagnosis and therapy of tumours,
 CC particularly cancer diagnosis. The present sequence is a protein
 CC used in the exemplification of the invention.

SO Sequence 137 AA;

Query Match 100.0%; Score 80; DB 22; Length 137;
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGGSTNYSALM 15
 |||||||
 DB 69 vimaggstnysalm 83

RESULT 12

AAR32843
 ID AAR32843 standard; Protein; 222 AA.

XX AAR32843;

XX

DT 19-JUN-1993 (first entry)
 XX
 DE VH NQ2/12.4-VK NQ10/12.5 linked peptide sequences #2.

KW Primer: human; immunoglobulin; Ig; variable region; VH; VL; Ck; JH;
 KW lymphocyte; vector; soluble; antibody; phage; linker; back; VH3;
 KW nested; In-cell PCR; cloning; polymorphic; TCR V; antiphenylloxazone;
 KW hydridoma; NQ2/12.4; NQ10/12.5.

XX Synthetic.

XX Key Location/Qualifiers

FT Region 1..113

FT Peptide /Label= VH_NQ2/12.4

FT /note= "Linker peptide"

FT Region 116..222

FT /Label= Vkappa_NQ2/12.4

PN WO9303151-A.

PD 18-FEB-1993.

PE 10-AUG-1992; 92WO-GB01483.

PR 10-AUG-1991; 91GB-0017352.

PR 11-JUN-1992; 92GB-0012419.

PA (MEDICAL RES COUNCIL.

PI Embleton MJ, Gorochoy G, Jones PT, Winter GP;

DR WPI; 1993-076508/09.

DR N-PSDB; AAQ37462.

PT Treatment of cell populations, partic. hybridomas - to link
 PT together copies of 2 or more non-contiguous DNA sequences to
 PT facilitate analysis

XX Disclosure; fig 5; 72pp; English.

XX The sequences given in AAR32840-43 show the mature heavy chain VH
 CC domains and the VK light chain genes of the antiphenylloxazone
 CC hybridomas NQ2/12.4 and NQ10/12.5 which have been linked via a linker
 CC peptide by In-cell PCR. The cDNA encoding these peptides was the JH
 CC segment, followed by assembly with linker primers, VH back primers
 CC based on the VH3 leader sequence and a forward Ck primer nested in
 CC the cells is then amplified with nested primers annealing to the 5',
 CC end of the VH gene and the 3' end of the Jk segment. In-cell PCR may
 CC be used to determine gene linkage analysis, particularly for the
 CC cloning of gene combinations that are polymorphic within a population
 CC of cells, such as the rearranged genes for Ig or TCR V regions.

SO Sequence 222 AA;

Query Match 100.0%; Score 80; DB 14; Length 222;
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGGSTNYSALM 15
 |||||||
 DB 50 vimaggstnysalm 64

RESULT 13

AAR32840
 ID AAR32840 standard; Protein; 235 AA.

XX AAR32840;

XX

DT	19-JUN-1993	(first entry)
XX		
DE	VH NQ2/12.4-VK NQ10/12.5	linked peptide sequences.
XX		
KW	Primer: human; immunoglobulin; Ig; variable region; VH; VL; Ck; JH;	
KW	lymphocyte; vector; soluble; antibody; phage; linker; back; VH3;	
KW	nested; In-cell PCR; cloning; polymorphic; TCR V; antiphenylloxazalone;	
KW	hybridoma; NQ2/12.4; NQ10/12.5.	
XX		
OS	Synthetic.	
XX		
EH	Key	Location/Qualifiers
FT	Region	1..113
FT		/label=VH_NQ2/12.4
FT	Peptide	114..127
FT		/note="Linker peptide"
FT	Region	128..235
FT		/label=VKappa_NQ2/12.4
PN		
XX	MO9303151-A.	
XX		
PD	18-FEB-1993.	
XX		
PF	10-AUG-1992;	92WO-GB01483.
XX		
PR	10-AUG-1991;	91GB-0017352.
PR	11-JUN-1992;	92GB-0012419.
XX		
PA	(MEDI-) MEDICAL RES COUNCIL.	
XX		
PL	Embleton MJ, Gorochov G, Jones PT, Winter GP;	
XX		
DR	WPI; 1993-076508/09.	
DR	N-PSDB; AAQ37459.	
XX		
PT	Treatment of cell populations, partic. hybridomas - to link	
PT	together copies of 2 or more non-contiguous DNA sequences to	
PT	facilitate analysis	
XX		
PS	Disclosure; Fig 2; 72pp; English.	
XX		
CC	The sequences given in AAR32840-43 show the mature heavy chain VH	
CC	domains and the VK light chain genes of the antiphenylloxazalone	
CC	hybridomas NQ2/12.4 and NQ10/12.5 which have been linked via a linker	
CC	peptide by in-cell PCR. The cDNA encoding these peptides was	
CC	synthesised using forward primers annealing to the Ck gene and the JH	
CC	segment, followed by assembly with linker primers, VH back primers	
CC	based on the VH3 leader sequence and a forward Ck primer nested in	
CC	respect to the primer used for cDNA. The assembled product within	
CC	the cells is then amplified with nested primers annealing to the 5'	
CC	end of the VH gene and the 3' end of the JK segment. In-cell PCR may	
CC	be used to determine gene linkage analysis, particularly for the	
CC	cloning of gene combinations that are polymorphic within a population	
CC	of cells, such as the rearranged genes for Ig or TCR V regions.	
XX		
SO	Sequence	235 AA:
XX		
QY	Query Match	100.0%; Score 80; DB 14; Length 235;
XX	Best Local Similarity	100.0%; Pred. No. 2.6e-05;
XX	Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 VIMAGGSTRYNSALM 15	
XX		
DB	50 VIMAGGSTRYNSALM 64	
XX		
RESULT 14		
AAR32569		
ID	AAR32569	standard; Protein: 269 AA.
XX		
AC	AAR32569;	
XX		

```

DT      08-JUN-1993   (first entry)
XX
DE      Fusion protein encoded by Ox VH-hinge-VL insert.
XX
XX      Spacer peptide; secretable; single chain; antibody; recombinant;
KW      scab; rDNA; linker; Bos taurus.
XX
OS      Synthetic.
XX
FH      Key          Location/Qualifiers
FT      Region       1..22
FT                      /note= "SS"
FT      Region       23..135
FT                      /note= "VH"
FT      Region       136..163
FT                      /note= "CBHI hinge"
FT      Region       164..277
FT                      /note= "VL"
XX
PN      FI9103434-A.
PN      WO9302198-A. (First Major Country Equivalent).
XX
PD      17-JAN-1992.
XX
PF      16-JUL-1991;    91WO-0913434.
XX
PR      16-JUL-1990;    90US-0552751.
XX
PA      (TERR-) TECH RES CENT FINLAND.
XX
PI      Alftan K, Knowles JKC, Laukkanen ML, Sizmann D, Takkinen K;
PI      Teeri TT;
XX
DR      WPI: 1992-134225/17.
DR      N-PSTDB: AAO36982.
XX
PT      Prod. of single chain fusion protein, pref. antibody - comprises
PT      transforming host cells, e.g. E. coli with expression constructs
PT      composed of proteins or domains, linked by spacer peptide(s)
XX
PS      Example; Fig 4; 56pp; English.
XX
CC      The sequence is that of the fusion protein encoded by the Ox
CC      VH-CBH1 hinge-VL insert which was used as part of a method for
CC      cloning secretable, biologically active single chain antibodies
CC      (scabs) and other secretable fusion proteins having at least 2
CC      distinct functional proteins or domains.
XX
SQ      Sequence      269 AA;

Query Match           100.0%; Score 80; DB 13; Length 269;
Best Local Similarity 100.0%; Pred. No. 3.le-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VIWAGGSTNYNSALM 15
        |||iiiiiiiii|||
Db       72 viwaggstnynsaln 86

RESULT 15
AABB1972
ID      AABB1972 standard; Protein; 561 AA.
XX
AC      AABB1972;
XX
DE      03-JUL-2001 (first entry)
XX
DE      Ganglioside GD2 specific antibody related protein SEQ ID NO: 31.
XX
KW      Ganglioside; GD2; complementation determining region; CDR; antibody;
KW      mouse; cancer.
XX
```

OS Synthetic.
 XX
 PN WO200123573-A1.
 XX
 PD 05-APR-2001.
 XX
 PE 29-SEP-2000; 2000MO-JP06773.
 XX
 PR 30-SEP-1999; 99JP-0278290.
 XX
 PA (KYOW) KYOWA HAKKO KOCYO KK.
 XX
 PI Hanai N, Shitara K, Nakamura K, Niwa R;
 DR WPI; 2001-266163/27.
 XX
 PT Human type complementation-determining domain transplanted antibody and
 PT derivatives against ganglioside GD2, useful in diagnosis and therapy of
 PT e.g. tumours, has low antigenicity, little side effects but potent
 PT activity in cancer
 XX
 PS Example 3; Page 111-114; 123pp; Japanese.
 XX
 CC The present invention describes an antibody, which can react specifically
 CC with ganglioside GD2, and is transplanted with a human type
 CC complementation-determining domain (CDR), or its fragments. The antibody
 CC and its derivatives are useful in diagnosis and therapy of tumours,
 CC particularly cancer diagnosis. The present sequence is a protein
 CC used in the exemplification of the invention.
 XX
 SQ Sequence 581 AA;

Query Match 100.0%; Score 80; DB 22; Length 581;
 Best Local Similarity 100.0%; Pred. No. 7.4e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VIMAGSTNYSALM 15
 ||||||||||||
 Db 50 VIMAGSTNYSALM 64

Search completed: August 20, 2002, 13:15:10
 Job time: 104 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:16:02 ; Search time 46.28 Seconds
(Without alignments)
31.144 Million cell updates/sec

Title: US-09-824-286-3_COPY_47_61
Perfect score: 80
Sequence: 1 VIMAGSTNYSALM 15

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	100.0	95	2 S17605	Ig heavy chain V r
2	80	100.0	100	2 S14490	Ig heavy chain V r
3	80	100.0	106	2 S26322	Ig heavy chain V r
4	80	100.0	106	2 S14489	Ig heavy chain V r
5	80	100.0	107	2 S14492	Ig heavy chain V r
6	80	100.0	107	2 S14493	Ig heavy chain V r
7	80	100.0	109	2 S11109	Ig heavy chain V r
8	80	100.0	112	2 S11100	Ig heavy chain V r
9	80	100.0	112	2 S11108	Ig heavy chain V r
10	80	100.0	113	2 S11101	Ig heavy chain V r
11	80	100.0	114	2 S11099	Ig heavy chain V r
12	80	100.0	114	2 S11106	Ig heavy chain V r
13	80	100.0	115	2 S11103	Ig heavy chain V r
14	80	100.0	116	2 S11102	Ig heavy chain V r
15	80	100.0	117	2 S10111	Ig heavy chain V r
16	80	100.0	120	2 PL0087	Ig heavy chain V r
17	80	100.0	141	2 S52446	Ig heavy chain V r
18	77	96.2	139	2 A32456	Ig heavy chain pre
19	76	95.0	109	2 PH1025	Ig heavy chain V r
20	76	95.0	114	2 S11104	Ig heavy chain V r
21	76	95.0	114	2 S11105	Ig heavy chain V r
22	75	93.8	112	2 S11098	Ig heavy chain V r
23	71	88.8	107	2 S14491	Ig heavy chain V r
24	70	87.5	118	2 P00266	Ig heavy chain V r
25	69	86.2	140	2 S55028	Ig heavy chain V r
26	67	83.8	116	2 S42484	Ig heavy chain V r
27	67	83.8	118	2 S32786	Ig heavy chain (an
28	66	82.5	231	2 PC4155	Ig gamma-2b chain
29	63	78.8	107	2 S14506	Ig heavy chain V r

30	63	78.8	121	2 S33131	Ig heavy chain V r
31	62	77.5	110	2 PH1024	Ig heavy chain V r
32	61	76.2	90	2 A49042	Ig heavy chain V r
33	61	76.2	100	2 A25913	Ig heavy chain pre
34	61	76.2	115	2 S26470	Ig heavy chain V r
35	61	76.2	115	2 S11107	Ig heavy chain V r
36	61	76.2	116	1 G1MS10	Ig heavy chain pre
37	61	76.2	116	2 A33932	Ig mu chain precu
38	61	76.2	127	2 B31807	Ig heavy chain V r
39	61	76.2	135	2 S31913	Ig gamma-2a chain
40	61	76.2	140	2 S14238	Ig gamma-1 chain p
41	60	75.0	117	2 S38563	Ig heavy chain V r
42	60	75.0	121	2 D30560	Ig heavy chain V r
43	60	75.0	122	2 A49049	Ig heavy chain V r
44	59	73.8	97	2 S55372	Ig heavy chain V r
45	59	73.8	101	2 S03466	Ig heavy chain V r

ALIGNMENTS

RESULT 1
S17605
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 06-Jun-1997
C:Accession: S17605
R:Jackson, T.; Hoogenboom, H.R.; Griffiths, A.D.; Winter, G.
Nature 352, 624-628, 1991
A:Title: Making antibody fragments using phage display libraries.
A:Reference number: S17230; MUID:91326098
A:Accession: S17605
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-95 <CLAS>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:8-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 3.9e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYSALM 15
DB 43 VIMAGSTNYSALM 57

RESULT 2

S14490
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C:Accession: S14490
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.
submitted to the FMBL data library, March 1991
A:Description: Natural polyclonal antibodies differ from Ag-induced antibodies in V
A:Reference number: S14484
A:Accession: S14490
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-100 <CHES>
A:Cross-References: EMBL:X58647; NID:951283; PIDN:CAA41504.1; PID:951284
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 4.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15
|||||
Db 50 VIMAGSTNYSALM 64

RESULT 3

S25322
Ig heavy chain V region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 19-Mar-1998 #sequence_revision 19-Mar-1998 #text_change 21-Jan-2000
C:Accession: S25322
R:Stark, S.E.; Caton, A.J.
J. Exp. Med. 174, 613-624, 1991
A:Title: Antibodies that are specific for a single amino acid interchange in a protein
A:Reference number: S25309; MUID:91341421
A:Accession: S25322
A:Molecule type: mRNA
A:Residues: 1-106 <STRA>
A:Cross-references: EMBL:X59182
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: heterotetramer; Immunoglobulin
F:15-91/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15
|||||
Db 44 VIMAGSTNYSALM 58

RESULT 4

S14489
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C:Accession: S14489
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.
submitted to the EMBL Data Library, March 1991
A:Description: Natural polypeptide antibodies differ from Ag-induced antibodies in VH
A:Reference number: S14484
A:Accession: S14489
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-106 <CHE>
A:Cross-references: EMBL:X58646; NID:951281; PIDN:CAA41503.1; PID:951282
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: heterotetramer; Immunoglobulin
F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15
|||||
Db 50 VIMAGSTNYSALM 64

RESULT 5

S14492
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C:Accession: S14492
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.
submitted to the EMBL Data Library, March 1991
A:Description: Natural polypeptide antibodies differ from Ag-induced antibodies in VH
A:Reference number: S14484
A:Accession: S14492

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-107 <CHE>
A:Cross-references: EMBL:X58649; NID:951287; PIDN:CAA41506.1; PID:951288
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: heterotetramer; Immunoglobulin
F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15
|||||
Db 50 VIMAGSTNYSALM 64

RESULT 6

S14493
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C:Accession: S14493
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.
submitted to the EMBL Data Library, March 1991
A:Description: Natural polypeptide antibodies differ from Ag-induced antibodies in V
A:Reference number: S14484
A:Accession: S14493
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-107 <CHE>
A:Cross-references: EMBL:X58650; NID:951289; PIDN:CAA41507.1; PID:951290
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: heterotetramer; Immunoglobulin
F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15
|||||
Db 50 VIMAGSTNYSALM 64

RESULT 7

S11109
Ig heavy chain V region (clone NQ5-89.4) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 21-Jan-2000
C:Accession: S11109
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazol
A:Reference number: S07331; MUID:83271467
A:Accession: S11109
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-109 <KAA>
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
F:15-93/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 4.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15
|||||
Db 50 VIMAGSTNYSALM 64

```
RESULT 8
S11100
Ig heavy chain V region (clone NQ2-20.5.3) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000
C:Accession: S11100
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11100
A:Molecule type: mRNA
A:Residues: 1-112 <NAT>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 80; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 4.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
    |||||||
Db 50 VIMAGSTNYNSALM 64

RESULT 9
S11108
Ig heavy chain V region (clone NQ5-78.2.6) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000
C:Accession: S11108
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11108
A:Molecule type: mRNA
A:Residues: 1-112 <NAT>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 80; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 4.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
    |||||||
Db 50 VIMAGSTNYNSALM 64

RESULT 10
S11101
Ig heavy chain V region (clone NQ2-48.2.2) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 21-Jan-2000
C:Accession: S11101
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11101
A:Molecule type: mRNA
A:Residues: 1-113 <KAA>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 80; DB 2; Length 113;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
    |||||||
Db 50 VIMAGSTNYNSALM 64
```

```
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
    |||||||
Db 50 VIMAGSTNYNSALM 64

RESULT 11
S11099
Ig heavy chain V region (clone NQ2-17.4.1) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000
C:Accession: S11099
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11099
A:Molecule type: mRNA
A:Residues: 1-114 <NAT>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 80; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
    |||||||
Db 50 VIMAGSTNYNSALM 64

RESULT 12
S11106
Ig heavy chain V region (clone NQ5-96.2) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000
C:Accession: S11106
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11106
A:Molecule type: mRNA
A:Residues: 1-114 <NAT>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 80; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
    |||||||
Db 50 VIMAGSTNYNSALM 64

RESULT 13
S11103
Ig heavy chain V region (clone NQ5-61.1.2) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000
C:Accession: S11103
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11103
A:Molecule type: mRNA
A:Residues: 1-115 <NAT>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>
```

F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 115;

Best Local Similarity 100.0%; Pred. No. 4.7e-06; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15
|||||
Db 50 VIMAGSTNYNSALM 64

RESULT 14

S11102

Ig heavy chain V region (clone NQ5-4.3.1) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000

C:Accession: S11102

R:Kaartinen, M.; Giffiths, G.M.; Markham, A.F.; Milstein, C.

Nature 304, 320-324, 1983

A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone

A:Reference number: S07331; MUID:83271467

A:Accession: S11102

A:Molecule type: mRNA

A:Residues: 1-116 <NAT>

C:Superfamily: Immunoglobulin V region; immunoglobulin homology

F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 116;
Best Local Similarity 100.0%; Pred. No. 4.8e-06; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15
|||||
Db 50 VIMAGSTNYNSALM 64

RESULT 15

S10111

Ig heavy chain V region (clone 26) precursor - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 23-Jul-1999

C:Accession: S10111

R:Kaartinen, M.; Solin, M.L.; Maekela, O.

EMBO J. 8, 1743-1748, 1989

A:Title: 'Allelic' forms of immunoglobulin V genes in different strains of mice.

A:Reference number: S10111; MUID:89356648

A:Accession: S10111

A:Molecule type: mRNA

A:Residues: 1-117 <KAA>

A:Cross-references: EMBL:X15471; NID:g50005; PIDN:CAA33499.1; PID:g50006

C:Superfamily: Immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:34-116/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.8e-06; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15
|||||
Db 69 VIMAGSTNYNSALM 83

Search completed: August 20, 2002, 13:16:02
Job time: 156 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:51 ; Search time 21.48 Seconds

(without alignments)
27.039 Million cell updates/sec

Title: US-09-824-286-3_COPY_47_61

Perfect score: 80
Sequence: 1 VIMAGSTNYNSALM 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	76.2	116	1 HV45_MOUSE	P01821 mus musculus
2	56	70.0	115	1 HV44_MOUSE	P01820 mus musculus
3	55	68.8	507	1 YG46_YEAST	P53301 saccharomyc
4	52	65.0	144	1 HV43_MOUSE	P01819 mus musculus
5	43	53.8	317	1 P1MT_THEMA	O56303 thermotoga
6	40	50.0	311	1 YD17_SCHPO	P87055 schistosach
7	40	50.0	318	1 NUI1M_ATEPA	O78653 atelies panl
8	40	50.0	318	1 NUI1M_CEBAP	O78655 hombo sapien
9	40	50.0	318	1 NUI1M_HUMAN	O96126 hylobates 1
10	40	50.0	318	1 NUI1M_HYLLA	O92444 pongo pygma
11	40	50.0	318	1 NUI1M_PAPHA	P92718 pongo pygma
12	40	50.0	318	1 NUI1M_PONPA	P92650 pongo pygma
13	40	50.0	318	1 NUI1M_PONPP	O92444 pongo pygma
14	40	50.0	318	1 NUI1M_SAGOE	O78654 saguinus oe
15	39	48.8	402	1 OPS4_CANAL	P46556 candida alb
16	39	48.8	440	1 DCO_DROME	O76324 drosophila
17	38	47.5	318	1 NUI1M_TRIUV	O78707 trichosurus
18	38	47.5	321	1 NUI1M_ALIMI	O47868 alligator m
19	38	47.5	322	1 NUI1M_PELSU	O79670 pelomedusa
20	38	47.5	344	1 Y401_HUMAN	O43151 homo sapien
21	38	47.5	426	1 GUN2_ERWCH	P07103 erwilia chr
22	38	47.5	454	1 VE2_HPV37	O80903 human papil
23	38	47.5	486	1 ALIN_ALLSA	O01594 allium satl
24	38	47.5	575	1 YF8K_ECOLI	P76481 escherichia
25	38	47.5	1319	1 ACSC_ACEXY	P19450 acetobacter
26	38	47.5	2262	1 RRP_P12HT	P26676 human parai
27	37	46.2	146	1 HV21_HUMAN	P06331 homo sapien
28	37	46.2	103	1 MCEA_KLEPN	O92444 klebsiella
29	37	46.2	116	1 HV05_CARAU	P19181 carassius a
30	37	46.2	172	1 LECA_PLEMA	O00298 pleurodeles
31	37	46.2	265	1 CYSO_BUCAI	P57674 buchiera ap
32	37	46.2	283	1 LECA_SARPE	P05047 sarcophaga
33	37	46.2	456	1 HEMN_AOUAE	O67886 agutlex aeo

34	37	46.2	475	1 PPB_SERMA	P19147 serratia ma
35	37	46.2	476	1 LA14_LYCES	P29535 lycopersico
36	37	46.2	484	1 DCOR_NEUCR	P27121 neuropept
37	37	46.2	499	1 VG02_BRP22	P26745 bacterioph
38	37	46.2	604	1 AMYC_RH10R	P27683 rhizopus or
39	37	46.2	881	1 NAI1_PHAVU	P39865 phaseolus v
40	36	45.6	116	1 HV60_MOUSE	P18531 mus musculu
41	36	45.0	73	1 MALE_PHOLU	P41130 photorhabdu
42	36	45.0	121	1 HV01_MOUSE	P01745 mus musculu
43	36	45.0	230	1 P1MT_WHEAT	O43209 triticum ae
44	36	45.0	281	1 POT1_ECOLI	P31136 escherichia
45	36	45.0	301	1 MPT5_MYCLE	O05868 mycobacteri

ALIGNMENTS

RESULT 1					
HV45_MOUSE		STANDARD;	PRT;	116 AA.	
ID HV45_MOUSE					
AC P01821;					
DT 21-JUL-1986 (Rel. 01, Created)					
DT 21-JUL-1986 (Rel. 01, Last sequence update)					
DE 15-JUL-1999 (Rel. 38, Last annotation update)					
DE Ig heavy chain V region MC101 precursor.					
OS Mus musculus (Mouse).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murine; Mus.					
OX NCBI_TaxID=10090;					
RN [1]					
RP SEQUENCE FROM N.A.					
RX MEDLINE=82075900; PubMed=6273429;					
RA Katoaka T., Nikaido T., Miyata T., Moriaki K., Honjo T.;					
RT "The nucleotide sequences of rearranged and germline immunoglobulin					
RT VH genes of a mouse myeloma MC101 and evolution of VH genes in					
RT mouse."					
RL J. Biol. Chem. 257:277-285(1982).					
CC					
CC This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC the European Bioinformatics Institute. There are no restrictions on its					
CC use by non-profit institutions as long as its content is in no way					
CC modified and this statement is not removed. Usage by and for commercial					
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC or send an email to license@isb-sib.ch).					
CC					
DR EMBL; J00502; AAA38515.1; -					
DR PIR; A02096; GIMS10.					
DR InterPro; IPR003006; Ig_MHC.					
DR InterPro; IPR003596; Ig_V.					
DR Pfam; PF00047; Ig; 1.					
DR SMART; SM00406; IGV; 1.					
DR Immunoglobulin V region; Signal.					
FT SIGNAL					
FT CHAIN					
FT NON_TER					
FT SEQUENCE					
116 AA; 12593 MW; 807966E7C52B3E CRC64;					
Query Match					
Best Local Similarity					
Matches 10; Conservative					
Score 61; DB 1; Length 116;					
Pred. No. 0.0017;					
Mismatches 1; Indels 0; Gaps 0;					
OY 1 VIMAGSTNYNSALM 15					
DB 69 VIMSGSTDYNAFRT 83					
RESULT 2					
HV44_MOUSE		STANDARD;	PRT;	115 AA.	
ID HV44_MOUSE					
AC P01820;					
DT 21-JUL-1986 (Rel. 01, Created)					

```

DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain V region PJ14 precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81012133; PubMed=6774258;
RA Sakano H., Maki R., Kurosawa Y., Roeder W., Tonegawa S.;
RT "Two types of somatic recombination are necessary for the generation
RT of complete immunoglobulin heavy-chain genes.";
RL Nature 286:676-683(1980).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V00767; CAA24148.1; -.
DR PIR; A02095; HVM514.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; I9_V.
DR Pfam; PF00047; I9; 1.
DR SMART; SM00406; IGV; 1.
DR Immunoglobulin V region; Signal.
FT SIGNAL 1 19
FT CHAIN 20 115 IG HEAVY CHAIN V REGION PJ14.
FT NON_TER 115 115
SO SEQUENCE 115 AA; 12447 MW; 7569DD4A4843D500 CRC64;

Query Match 70.0%; Score 56; DB 1; Length 115;
Best Local Similarity 71.4%; Pred. No. 0.012;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSAL 14
Db 69 MIMGDSIDYNSAL 82

RESULT 3
Y64E_YEAST STANDARD; PRT; 507 AA.
AC P53301;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 52.8 kDa protein in BUB1-HIP1 intergenic region.
GN YGR189C OR G7553.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97279231; PubMed=9133739;
RA Arroyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M.,
RA Nombela C.;
RT "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm
RT of Saccharomyces cerevisiae chromosome VII.";
RL Yeast 13:357-363(1997).
CC -----
CC -1- SIMILARITY: SOME, TO YEAST UTR2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z72974; CAA97215.1; -.
DR EMBL; X99074; CAA67525.1; -.
DR HSSP; P23804; IAKK.
DR SGD; S0003421; CRH1.
DR InterPro; IPR00757; Glyco_hydro_16.
DR Pfam; PF00722; Glyco_hydro_16; 1.
KW Hypothetical protein.
FT DOMAIN 63 66 POLY-SER.
FT DOMAIN 301 310 POLY-SER.
FT DOMAIN 345 357 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 467 470 POLY-SER.
SO SEQUENCE 507 AA; 52757 MW; 7D7B61F57AE942C CRC64;

Query Match 68.8%; Score 55; DB 1; Length 507;
Best Local Similarity 81.8%; Pred. No. 0.081;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 WAGGSTNYNSA 13
Db 233 WAGGETNYNSA 243

RESULT 4
HY43_MOUSE STANDARD; PRT; 144 AA.
AC P01819;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain V region MOPC 141 precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81012133; PubMed=6774258;
RA Sakano H., Maki R., Kurosawa Y., Roeder W., Tonegawa S.;
RT "Two types of somatic recombination are necessary for the generation
RT of complete immunoglobulin heavy-chain genes.";
RL Nature 286:676-683(1980).
CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS TRANSLATED FROM A
CC DIFFERENTIATED GENE ISOLATED FROM A MYELOMA THAT SECRETES IGG2B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00491; AAA8121.1; -.
DR EMBL; V00768; CAA24149.1; -.
DR PIR; A02094; G2MS14.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; I9_V.
DR Pfam; PF00047; I9; 1.
DR SMART; SM00406; IGV; 1.
DR Immunoglobulin V region; Signal.
FT SIGNAL 1 19
FT CHAIN 20 144 IG HEAVY CHAIN V REGION MOPC 141.
FT NON_TER 144 144
SO SEQUENCE 144 AA; 15759 MW; 8E47A7CB3706D30A CRC64;

Query Match 65.0%; Score 52; DB 1; Length 144;

```

Best Local Similarity 69.2%; Pred. No. 0.069;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Oy 2 IWAGSTNYSAL 14
|||:|||||
Db 70 IWNGSTIDNSTL 82

RESULT 5
PIMT_THEME
ID PIMT_THEME STANDARD: PRT: 317 AA.
AC Q56308;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-L-isospartate O-methyltransferase (EC 2.1.1.77) (Protein-
beta-aspartate methyltransferase) (PIMT) (Protein L-isospartyl
methyltransferase) (L-isospartyl protein carboxyl methyltransferase).
GN PCM OR TM0704.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96134983; PubMed=8550470;
RA Swanson R.V., Sanna M.G., Simon M.I.;
RT "Thermostable chemotaxis proteins from the hyperthermophilic
bacterium Thermotoga maritima."
RL J. Bacteriol. 178:484-489(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
Hart D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Saizberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=9903468; PubMed=9784234;
RA Ichikawa J.K., Clarke S.;
RT "A highly active protein repair enzyme from an extreme thermophile:
the L-isospartyl methyltransferase from Thermotoga maritima."
RL Arch. Biochem. Biophys. 358:222-231(1998).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX PubMed=11080641;
RA Skinner M.M., Puvathingal J.M., Walter R.L., Friedman A.M.;
RT "Crystal structure of protein isospartyl methyltransferase. A
catalyst for protein repair."
RL Structure 8:1189-1201(2000).
CC -I- FUNCTION: CATALYZES THE METHYL ESTERIFICATION OF L-ISOSPARTYL
RESIDUES IN PEPTIDES AND PROTEINS THAT RESULT FROM SPONTANEOUS
DECOMPOSITION OF NORMAL L-ASPARTYL AND L-ASPARAGINYL RESIDUES. IT
PLAYS A ROLE IN THE REPAIR AND/OR DEGRADATION OF DAMAGED PROTEINS.
CC -I- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + protein L-beta-
aspartate = S-adenosyl-L-homocysteine + protein L-beta-aspartate
methyl ester.
CC -I- SUBUNIT: MONOMER.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- MISCELLANEOUS: EXTREMELY HEAT STABLE, WITH NO LOSS OF ACTIVITY
AFTER 60 MIN AT 100 DEGREES CELSIUS. ENZYME ACTIVITY IS OBSERVED
AT TEMPERATURES AS HIGH AS 93 DEGREES CELSIUS WITH AN OPTIMAL
ACTIVITY OF 164 NMOL/MIN/MG PROTEIN AT 85 DEGREES CELSIUS.
CC -I- SIMILARITY: BELONGS TO THE L-ISOSPARTYL/D-ASPARTYL PROTEIN
METHYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U30501; AAA96385.1; -
DR EMBL: AE001742; AAD35766.1; -
DR PDB: 1DL5; 08-DEC-00.
DR TIGR: TM0704; -
DR InterPro: IPR000682; PCMT.
DR InterPro: IPR000051; SAM_bind.
DR Pfam: PF01135; PCMT; 1.
DR PROSITE: PS01279; PCMT; 1.
KW Transferase; Methyltransferase; 3D-structure; Complete proteome.
SQ SEQUENCE 317 AA; 36400 MW; 2FE6019571ADDF2C CRC64;

Query Match 53.8%; Score 43; DB 1; Length 317;
Best Local Similarity 53.3%; Pred. No. 5.2;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Oy 1 VIMAGSTNYSALM 15
|: |||:|:|
Db 79 VLEIGGCTGYNAVM 93

RESULT 6
YD77_SCHPO
ID YD77_SCHPO STANDARD: PRT: 311 AA.
AC P87055;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 35.1 kDa protein C57A10.07 in chromosome I.
GN SPAC57A10.07.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Badcock K., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO YEAST YOR238W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z94864; CAB08170.1; -
KM Hypothetical protein; Transmembrane.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
SQ SEQUENCE 311 AA; 35128 MW; 6EE388FDE2DF3C64 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 311;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Oy 2 IWAGSTN 9
:|:|||||
Db 87 VWLGGSTN 94

```

RESULT 7
ID NUIM_ATEPA STANDARD; PRT: 318 AA.
AC 078693;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN MTND1 OR NDI.
OS Ateles paniscus (Black spider monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Platyrrhini; Cebidae; Atelinae; Ateles.
OX NCBI_TaxID=9510;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=98404151; PubMed=9732458;
RA Cao Y., Janke A., Weddell P.J., Westernman M., Takenaka O., Murata S.,
RA Okada N., Pabo S., Hasegawa M.;
RT "Conflict among individual mitochondrial proteins in resolving the
phylogeny of eutherian orders."
J. Mol. Evol. 47:307-322(1998).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB010971; BAA32096.1; -
DR InterPro; IPR001694; Resp_chain_NADH_DHL.
DR Pfam; PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEX1_NDI_1; 1.
DR PROSITE; PS00668; COMPLEX1_NDI_2; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 318 AA; 35815 MW; 39BECF5A20D415F6 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 318;
Best Local Similarity 40.0%; Pred. NO. 17;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0.

QY 1 VIMAGGSTNYSALM 15
:::|::|::|::|
Db 116 ILMGSMASNNVYLI 130

RESULT 8
NUIM_CEBAP STANDARD; PRT: 318 AA.
AC 078695;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN MTND1 OR NDI.
OS Cebus apella (Brown-capped capuchin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Platyrrhini; Cebidae; Cebinae; Cebus.
OX NCBI_TaxID=9515;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=98404151; PubMed=9732458;
RA Cao Y., Janke A., Weddell P.J., Westernman M., Takenaka O., Murata S.,
RA Okada N., Pabo S., Hasegawa M.;
RT "Conflict among individual mitochondrial proteins in resolving the

```

```

RT phylogeny of eutherian orders.";
RL J. Mol. Evol. 47:307-322(1998).
CC -I- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -I- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
-----
DR EMBL; AB010973; BAA32098.1; -
DR InterPro; IPRO01694; resp_chain_NADH_DH1.
DR Pfam; PF00146; NADHdh_1.
DR Prosite; PS00667; COMPLEXI_NDL_1; 1.
DR PROSITE; PS00668; COMPLEXI_NDL_2; 1.
RW Oxidoreductase; NAD: Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 318 AA; 35805 MW; DBBAV93FE59AE0DA CRC64;

Query Match          50.0%; Score 40; DB 1; Length 318;
Best Local Similarity 40.0%; Pred No. 17;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0.

QY      1 VIMAGSGTNYNSALM 15
         ::|::|::|::|::|
Db       116 ILMSGWSNYSNYALI 130

RESULT           9
NUO1M_HUMAN      STANDARD;             PRT;     318 AA.
ID   NUO1M_HUMAN
AC   P03886; Q37523;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN   MTND1 OR ND1.
OS   Homo sapiens (Human).
OC   Mitochondrion.
OC   Mammalia; Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC   Eukaryota; Eubacteria; Primates; Catarrhini; Homnidae; Homo.
CX   NCBI_TaxID=9606;
[1]
RN   SEQUENCE FROM N.A.
RP   MEDLINE=81173052; PubMed=7219534;
RA   Anderson S., Bankier A.T., Barrell B.G., de Brulin M.H.L.,
RA   Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A.,
RA   Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G.;
RT   "Sequence and organization of the human mitochondrial genome.";
RL   Nature 290:457-465(1981).
[2]
RN   SEQUENCE FROM N.A., AND VARIANTS ALA-87 AND ALA-168.
RP   TISSUE=Placenta;
RC   MEDLINE=95132634; PubMed=7530363;
RA   Horai S., Hayasaka K., Kondo R., Tsugane K., Takahata N.;
RA   "Recent African origin of modern humans revealed by complete sequences
RT   of hominoid mitochondrial DNAs.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:532-536(1995).
[3]
RN   SEQUENCE OF 130-318 FROM N.A.
RP   MEDLINE=81170577; PubMed=6260957;
RA   Sanger F., Coulson A.R., Barrell B.G., Smith A.J.H., Roe B.A.;
RT   "Cloning in single-stranded bacteriophage as an aid to rapid DNA
RT   sequencing.";
RL   J. Mol. Biol. 143:161-178(1980).
[4]
RN   IDENTIFICATION OF PROTEIN.
RP   MEDLINE=85188293; PubMed=3921850;
RA   Chomyn A., Mariotti P., Cleeter M.W.J., Ragan C.I., Matsuno-Yagi A.,
RA   Hateli Y., Doollittle R.F., Altardi G.;

```


GN OPS4.
OS *Candida albicans* (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WO-1;
 RX MEDLINE=9339284; PubMed=8478072;
 RA Morrow B., Srikantha T., Anderson J., Soli D.R.;
 RT "Coordinate regulation of two opaque-phase-specific genes during
 RL while-opaque switching in Candida albicans.";
 Intect. Immun. 61:1823-1828(1993).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L10735; -; NOT_ANNOTATED_CDS.
 KW Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 402 OPAQUE-PHASE-SPECIFIC PROTEIN OP4.
 FT DOMAIN 170 175 POLY-SER.
 FT DOMAIN 176 185 POLY-ALA.
 FT DOMAIN 341 349 SER-RICH.
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 402 AA; 41340 MW; 0EEC56D0937E9258 CRC64;

Query Match 48.8%; Score 39; DB 1; Length 402;
 Best Local Similarity 58.3%; Pred. No. 31;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 OY 2 IWAGSTNYNSA 13
 |||||: ||
 DB 165 IWLGGSSSSSA 176

Search completed: August 20, 2002, 13:17:51
 Job time: 260 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:23 ; Search time 73.7 Seconds

(without alignments)
35.209 Million cell updates/sec

Title: US-09-824-286-3_COPY_47_61

Perfect score: 80
Sequence: 1 VIMAGSTNYNSALM 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_19:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_protist:*
13: sp_virus:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriophage:*
17: sp_archaeophages:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	81.2	482	11	091X92
2	51	63.8	121	11	099NG4
3	46	57.5	937	5	017469
4	45	56.2	449	9	09AY20
5	44	55.0	328	16	092X08
6	44	55.0	613	4	096EX0
7	43	53.8	383	16	092N74
8	43	53.8	497	16	0927U5
9	43	53.8	1390	10	049145
10	42	52.5	145	11	092202
11	42	52.5	453	17	097XP6
12	42	52.5	528	11	09C084
13	42	52.5	600	3	042795
14	42	52.5	662	3	099012
15	42	52.5	696	16	09CHC2
16	41.3	51.9	597	4	09B010

17	41.5	51.9	618	4	096AA6
18	41	51.2	215	2	09JPN0
19	41	51.2	489	2	P72495
20	41	51.2	489	2	P72497
21	40	50.0	302	3	042800
22	40	50.0	318	8	09B1M4
23	40	50.0	318	8	09B1L5
24	40	50.0	318	8	09P9W3
25	40	50.0	318	8	09P9Z0
26	40	50.0	318	8	09T9X9
27	40	50.0	318	8	09T9X2
28	40	50.0	318	8	09B2Y9
29	40	50.0	318	8	09B2X5
30	40	50.0	318	8	09B2X4
31	40	50.0	318	8	09B2V7
32	40	50.0	318	8	09B2U3
33	40	50.0	318	8	09G393
34	40	50.0	318	8	094Z91
35	40	50.0	318	8	094P91
36	40	50.0	319	8	09TEH1
37	40	50.0	628	9	056785
38	40	50.0	824	10	0940D1
39	40	50.0	825	10	09SH62
40	40	50.0	1013	10	09LK23
41	40	50.0	1013	10	093Z55
42	40	50.0	1021	10	09LH88
43	40	50.0	1088	10	09C7C9
44	40	50.0	1194	2	09R6P0
45	40	50.0	1451	10	049140

ALIGNMENTS

RESULT 1
ID 091X92 PRELIMINARY: PRT: 482 AA.
AC 091X92;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
OS UNKNOWN (PROTEIN FOR MGC:18822).
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerogonathi; Muridae; Murinae; Mus.
OX NCBI_Taxid-10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON;
RA Strausberg R.;
DR Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC011181; AAH1181.1; -
SQ SEQUENCE 482 AA: 51665 MW: 312E01F9C1BC7F3C CRC64;

Query Match 81.2%; Score 65; DB 11; Length 482;
Best Local Similarity 85.7%; Pred. No. 0.0095;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 VIMAGSTNYNSAL 14
Db 69 VIMTGVNTYNSAL 82
RESULT 2
ID 099NG4 PRELIMINARY: PRT: 121 AA.
AC 099NG4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SINGLE CHAIN FV (FRAGMENT).
OS Mus musculus (Mouse).

OC Plasmid pHEM1.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BAIB/C.
RX MEDLINE=98169018; PubMed=9510199;
RA Haulisch H., Frank R., Hennecke M., Baensch M., Sohns B., Arseniev L.,
RA Bausch W., Kola A., Klos A., Koehl J.;
RT "Site-directed Cys-Receptor Antibodies from Phage Display Libraries";
RL J. Immunol. 160:2947-2958(1998).
DR EMBL, AJ222590; CAA10890.1; -.
DR HSSP; P01825; 7EAB.
DR InterPro; IPR003599; 1g.
DR InterPro; IPR003600; 1g_1like.
DR InterPro; IPR003006; 1g_MHC.
DR InterPro; IPR003596; 1g_v.
DR Pfam; PF00047; 1g_1.
DR SMART; SM00409; 1g; 1.
DR SMART; SM00406; 1g; 1.
DR SMART; SM00410; 1g_1like; 1.
KM Plasmid.
FT NON_TER
FT NON_TER
SC SEQUENCE 121 AA; 13255 MW; D293E4EBC8C59D5B CRC64;

Query Match 63.8%; Score 51; DB 11; Length 121;
Best Local Similarity 60.0%; Pred. No. 0.45;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15
ID 017469 PRELIMINARY; PRT; 937 AA.
AC 017469; O45065;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE OLFACTORY CHANNEL.
GN OSM-9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCB1_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97477445; PubMed=9334401;
RA Colbert H.A., Smith T.L., Bargmann C.I.;
RT "OSM-9, a novel protein with structural similarity to channels, is
RT required for olfaction, mechanosensation, and olfactory adaptation in
RT Caenorhabditis elegans.";
RL J. Neurosci. 17:8259-8269(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du Z., Magg J.L.;
RT "The sequence of C. elegans cosmid B0212.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF031408; AAB87064.1; -.

DR EMBL; AF045639; AAC02569.2; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000636; Cation_chan_non_11g.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR Pfam; PF00023; ank; 4.
DR Pfam; PF00520; ion_trans; 1.
DR SMART; SM00248; ANK; 2.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
KM ANK repeat; Repeat.
SC SEQUENCE 937 AA; 106847 MW; 1798CB866E814A3A CRC64;

Query Match 57.5%; Score 46; DB 5; Length 937;
Best Local Similarity 90.0%; Pred. No. 32;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 GSTNYNSALM 15
ID 09AYZ0 PRELIMINARY; PRT; 449 AA.
AC 09AYZ0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DNA TRANSFER PROTEIN.
GN 20.
OS Bacteriophage HK620.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCB1_TaxID=155148;
RN [1]
RP SEQUENCE FROM N.A.
RA Clark A.J., Inwood W.B., Cloutier T., Dhillon T.S.;
RT "Nucleotide Sequence of Coliphage HK620 and the Evolution of Lambdoid
RT Phages.";
RL J. Mol. Biol. 0:0-0(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Clark A.J., Inwood W.B., Cloutier T., Dhillon T.S.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF335358; AAK28901.1; -.
SQ SEQUENCE 449 AA; 48903 MW; D65283FD2860B7AB CRC64;

Query Match 56.2%; Score 45; DB 9; Length 449;
Best Local Similarity 53.8%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 3 WAGSTNYNSALM 15
ID 092X08 PRELIMINARY; PRT; 328 AA.
AC 092X08;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE TRANSCRIPTIONAL REGULATOR PROTEIN.
GN SMB20154.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCB1_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.
SPRAIN-1021:
RX MEDLINE=21368234; PubMed=11474104;
RA Gilbert F., Finan T.M., Long S.R.,
RA Barlow-Hubler F., Barnett M.J., Becker A., Bolstad P., Bothe G.,
RA Bouly M., Bowser L., Burnmaster J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Drenth N.A., Fisher R.F.,

RESULT	9	
049145	PRELIMINARY;	PRT; 1390 AA.
ID 049145		
AC 049145;		
DT 01-JUN-1998	(TREMBLrel. 06, Created)	
DT 01-JUN-1998	(TREMBLrel. 06, Last sequence update)	

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE POLYPROTEIN (FRAGMENT).
 OS Arabidopsis arenosa.
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eucotids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=38785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CARE-4; TRANSPOSON-EVELEKNEVEL RETROTRANSPOSON;
 RX MEDLINE=98250692; PubMed=9584105;
 RA Henikoff S., Comai L.;
 RT "A DNA methyltransferase homolog with a chromodomain exists in
 RL multiple polymorphic forms in Arabidopsis.";
 RT Genetics 149:307-318(1998).
 DR EMBL: AF039376; AAC02672.1; -
 DR MEROPS: A11.004;
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF00665; rve; 1.
 DR Pfam: PF00098; zf-CCHC; 1.
 DR SMART: SM00343; Znf_C2HC; 1.
 KW Zinc-finger.
 FT NON_TER
 SQ SEQUENCE 1390 AA; 154000 MW; FC37187A4FAA05CF CRC64;

Query Match 53.8%; Score 43; DB 10; Length 1390;
 Best Local Similarity 70.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 WAGSTNTNS 12
 Db 1298 WAGDTDNYS 1307

RESULT 10
 ID 092202 PRELIMINARY; PRT; 145 AA.
 AC 092202;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE RIKEN CDNA 5430416A05 GENE.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC006884; AA06884.1; -
 SQ SEQUENCE 145 AA; 16031 MW; 492EB74102B475BC CRC64;

Query Match 52.5%; Score 42; DB 11; Length 145;
 Best Local Similarity 47.6%; Pred. No. 18;
 Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

OY 3 WAGGS-----TNYSALM 15
 Db 84 WNGGSNYQANROTSYNSAKM 104

RESULT 11
 ID 097XP6 PRELIMINARY; PRT; 453 AA.
 AC 097XP6;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE CYTOSINE PERMEASE (CODB).

GN CODB.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aweyer M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moers A., Erasmo G., Fletcher C., Gordon P.M.K.,
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus p2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 DR EMBL: AB006780; AAK41877.1; -
 DR InterPro: IPR001248; Transp_cyl_pur.
 KW Complete proteome.
 SQ SEQUENCE 453 AA; 48816 MW; EF43F53B35471585 CRC64;

Query Match 52.5%; Score 42; DB 17; Length 453;
 Best Local Similarity 70.0%; Pred. No. 67;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VIMAGSTNY 10
 Db 43 VIMAGGALGX 52

RESULT 12
 ID 09CU84 PRELIMINARY; PRT; 528 AA.
 AC 09CU84;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE 5430416A05RIK PROTEIN (FRAGMENT).
 GN 5430416A05RIK.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=HEAD;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsunk S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK017312; BAB30687.1; -
 DR MGD: MGI:1918590; 5430416A05RIK.
 DR InterPro: IPR000687; RIO1_ZK632_MJ0444.
 DR InterPro: IPR001245; Tyr_pkinase.

DR Pfam; PF01163; RIO1; 1.
 DR SMART; SM00090; RIO; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
 FT NON_TER 528 528
 SQ SEQUENCE 528 AA; 60184 MW; 4AED139E966CD122 CRC64;

Query Match 52.5%; Score 42; DB 11; Length 528;
 Best Local Similarity 47.6%; Pred. No. 79;
 Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

OY 3 WAGS-----TNSALM 15
 :|||:|||||
 Db 84 WNGSGNYQANROTSNYSAMK 104

RESULT 13
 ID 042795 PRELIMINARY; PRT; 600 AA.
 AC 042795;
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE SERINE/THREONINE PROTEIN KINASE.
 GN LAPK.
 OS Colletotrichum trifolii.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetes; incertae sedis; Phyllachorales; Phyllachoraceae;
 OC mitosporic Phyllachoraceae; Colletotrichum.
 OX NCBI_TaxID=5466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yang Z., Dickman M.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AF046923; AAC04357.1; -.
 DR HSP; P05132; ICTP.
 DR InterPro; IPR000961; Euk_Pkinase.
 DR InterPro; IPR000961; Pkinase_C.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; Pkinase_2.
 DR Pfam; PF00433; Pkinase_C; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00133; S_TKc; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 600 AA; 66893 MW; C336051D3AE5AA6 CRC64;

Query Match 52.5%; Score 42; DB 3; Length 600;
 Best Local Similarity 66.7%; Pred. No. 92;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 IWAGSTNY 10
 :|||:|
 Db 132 LWAGSTNY 140

RESULT 14
 ID 099012 PRELIMINARY; PRT; 662 AA.
 AC 099012;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE SERINE/THREONINE PROTEIN KINASE.
 GN PKT1.
 OS Trichoderma reesei (Hypocrea jecorina).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Hypocreaceae; Hypocrea.
 OX NCBI_TaxID=51453;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-QW9414;
 RX MEDLINE=94357455; PubMed=8076837;
 RA Morawetz R., Mischak H., Goodnight J., Lendenfeld T., Mushinsky J.F.,
 RA Kudrick C.P.;
 RT "A protein kinase-encoding gene, pkl1, from Trichoderma reesei,
 RT homologous to the yeast YPK1 and YPK2 (YKR2) genes."
 RL Gene 146:309-310(1994).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; U05811; AAA57318.1; -.
 DR HSP; P05132; ICTP.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR000961; Pkinase_C.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; Pkinase_1.
 DR Pfam; PF00433; Pkinase_C; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00133; S_TKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 662 AA; 73380 MW; A33BDCD93FC24C23 CRC64;

Query Match 52.5%; Score 42; DB 3; Length 662;
 Best Local Similarity 66.7%; Pred. No. 1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 IWAGSTNY 10
 :|||:|
 Db 204 LWAGSTNY 212

RESULT 15
 ID 09CHC2 PRELIMINARY; PRT; 696 AA.
 AC 09CHC2;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
 DT 01-OCT-2001 (TREMblrel. 18, last annotation update)
 DE UNKNOWN PROTEIN.
 GN Y1CA.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Lactococcus.
 OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,
 RX MEDLINE=21235186; PubMed=11337471;
 RA Weissbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 RT lactis ssp. lactis IL1403."
 RL Genome Res. 11:731-753(2001).
 DR EMBL; AE006314; AAK04908.1; -.
 KW Complete proteome.
 SQ SEQUENCE 696 AA; 78889 MW; 52AD11A0A2B11EAC CRC64;

Query Match 52.5%; Score 42; DB 16; Length 696;
 Best Local Similarity 60.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 V1WAGSTNY 10
 :|:|:|
 Db 148 V1WAGSTNY 157

Search completed: August 20, 2002, 13:17:25
 Job time: 239 sec

PR 10-MAY-1996; 96US-0017466.
XX (BIOJ) BIOGEN INC.
XX Benjamin CD, Burkly LC, Hession C, Whitty A;
XX WPI: 1998-008885/01.
DR N-PSDB; AAT97441.
PR Blocking agents of the gamma common chain of cytokine receptors -
XX particularly monoclonal antibodies, used to induce T cell anergy for
XX treatment of immunological diseases
XX
XX Claim 23; Page 81-82; 111pp; English.
XX
XX This polypeptide comprises the heavy chain variable region (VH) of
XX monoclonal antibody (Mab) C9.B8, which is produced by a hybridoma
XX deposited as ATCC 12107, and which is specific for the gamma
XX constant (gc) chain (see AAW31646) of human cytokine receptors. The
XX invention provides compositions and methods for inhibiting cytokine
XX signalling using gc chain blocking agents for the treatment of
XX immunological diseases such as myasthenia gravis, rheumatoid
XX arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,
XX inflammatory bowel disease, sympathetic ophthalmia, uveitis,
XX allergy, asthma, parasitic infection, graft vs. host disease or
XX psoriasis. Preferred gc blocking agents include Mab C9.B8, its Fab
XX fragment and an antibody having a light chain variable region
XX CDR selected from those of CP.B8 VH or a heavy chain variable
XX region CDR selected from those of CP.B8 VL (see AAW31647).
SQ Sequence 112 AA;

Query Match 100.0%; Score 52; DB 19; Length 112;
Best Local Similarity 100.0%; Pred. No. 0.0099;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
Db 95 egstvdsmdy 104
|||||

RESULT 2
AAW80726
ID AAW80726 standard; Protein: 201 AA.
XX
XX AAW80726;
XX
XX 24-DEC-1998 (first entry)
XX
XX S. pneumoniae prolipoprotein diacylglyceryl transferase.
XX
XX Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
XX virulence; antibody; infection; detection; treatment; hypothetical;
XX cell wall biosynthetic; external target; minimal gene set protein.
XX
XX Streptococcus pneumoniae.
XX
XX WO9826072-A1.
XX
XX 18-JUN-1998.
XX
XX 09-DEC-1997; 97WO-US22578.
XX
XX 13-DEC-1996; 96US-0036281.
XX
XX (ELIL) LILLY & CO ELI.
XX
XX Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,
XX Mills BJ, Norris FH, Peery RB, Rockey PK, Rosteck PR,
XX Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ;
XX Young Bellido ML;
XX

DR WPI: 1998-348529/30.
DR N-PSDB; AAV65284.
XX
XX Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
XX for evaluating gene expression, and identification of virulence
XX genes
XX
XX Claim 3; Page 318; 333pp; English.
XX
XX This sequence represents a S. pneumoniae prolipoprotein diacylglyceryl
XX transferase. The invention provides DNA sequences (AAV65201 to AAV65304)
XX from the Streptococcus pneumoniae genome and corresponding protein
XX sequences (AAW80605 to AAW80728). The protein sequences are classified as
XX hypothetical, cell wall biosynthetic, external target, or minimal gene
XX set proteins. A recombinant host containing a vector comprising any of
XX the above nucleic acids can be used for the recombinant expression of the
XX proteins. The invention also provides a DNA chip having arrayed on it at
XX least 15 base pair fragment of any one or more of these DNA sequences.
XX The DNA chip can be used methods for evaluating gene expression in S.
XX pneumoniae and for identifying virulence genes in S. pneumoniae.
XX Antibodies that selectively bind to the above proteins or peptide
XX fragments can be used to treat S. pneumoniae infection. The antibodies
XX can also be used to detect S. pneumoniae cells.
SQ Sequence 201 AA;

Query Match 73.1%; Score 38; DB 19; Length 201;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSTVDSMDY 10
Db 81 gatvdnldy 89
|:||||:

RESULT 3
AAW38541
ID AAW38541 standard; Protein: 266 AA.
XX
XX AAW38541;
XX
XX 06-NOV-1998 (first entry)
XX
XX S. pneumoniae prolipoprotein diacylglyceryl transferase.
XX
XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;
XX immunological response; inoculation; antibody production; inhibitor;
XX T cell immune response; antimicrobial compound; bacterial adhesion;
XX extracellular matrix protein; protein-mediated cell invasion; wound;
XX pathogenesis.
XX
XX Streptococcus pneumoniae.
XX
XX WO9743303-A1.
XX
XX 20-NOV-1997.
XX
XX 14-MAY-1997; 97WO-US07950.
XX
XX 14-MAY-1996; 96US-0017670.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX (SMIK) SMITHKLINE BEECHAM PLC.
XX
XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
XX Stodola RK;
XX
XX WPI: 1998-008793/01.
XX N-PSDB; AAT98597.
XX
XX Novel Streptococcus pneumoniae proteins and related DNA - useful for
XX diagnosing anti-microbial agents for treatment of bacterial
XX

PT infections
 XX
 PS Claim 12: Pages 319-320; 483pp; English.
 XX
 CC This sequence represents a Streptococcus pneumoniae protein that, based
 CC on homology with a S. typhimurium protein, is a prolipoprotein
 CC diacylglycerol transferase, and is encoded by a DNA of the invention.
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 XX
 SQ Sequence 266 AA;
 Query Match 73.1%; Score 38; DB 19; Length 266;
 Best Local Similarity 66.7%; Pred. No. 16;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GSTVDSMDY 10
 Db 146 gatvdndly 154
 RESULT 4
 AAW83009 standard; Protein: 266 AA.
 XX
 AC AAW83009;
 DT 01-FEB-1999 (first entry)
 XX
 DE Streptococcus pneumoniae Lgt protein.
 DE
 KW Streptococcus pneumoniae; Lgt; pneumococcal; diagnosis; infection;
 KW otitis media; conjunctivitis; pneumonia; bacteraemia; meningitis;
 KW sinusitis; pleural empyema; endocarditis.
 XX
 OS Streptococcus pneumoniae.
 OS
 PN US5840538-A.
 XX
 PD 24-NOV-1998.
 XX
 PF 06-AUG-1997; 97US-0906753.
 XX
 PR 06-AUG-1997; 97US-0906753.
 PA (SMIR) SMITHKLINE BEECHAM CORP.
 PA
 PI Petit CM;
 XX
 DR WPI: 1999-034037/03.
 DR N-PSDB; AAV69831.
 XX
 PT Polynucleotide encoding pneumococcal Lgt polypeptide - used
 PT therapeutically and in diagnosis
 XX
 PS Claim 12; Column 21-24; 14pp; English.
 CC The present sequence represents a Streptococcus pneumoniae Lgt

CC protein, where the Streptococcus pneumoniae is of the deposited
 CC strain (NCIMB 40794). A host cell containing a vector comprising the
 CC nucleic acid sequence encoding the lgt protein, can be cultured to
 CC produce the protein, which is a Streptococcus pneumoniae lgt gene
 CC product. The nucleic acid sequence encoding the lgt protein can be
 CC used biologically, diagnostically, prophylactically, clinically or
 CC therapeutically. The types of diseases which can be diagnosed and
 CC treated are: otitis media; conjunctivitis; pneumonia; bacteraemia;
 CC meningitis; sinusitis; pleural empyema; and endocarditis.
 XX
 SQ Sequence 266 AA;
 Query Match 73.1%; Score 38; DB 20; Length 266;
 Best Local Similarity 66.7%; Pred. No. 16;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GSTVDSMDY 10
 Db 146 gatvdndly 154
 RESULT 5
 AAY81550 standard; Protein: 267 AA.
 XX
 AC AAY81550;
 DT 24-MAY-2000 (first entry)
 XX
 DE Streptococcus pneumoniae type 4 protein sequence #50.
 DE
 KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
 KW pneumococcal disease.
 XX
 OS Streptococcus pneumoniae.
 OS
 PN WO200006737-A2.
 XX
 PD 10-FEB-2000.
 XX
 PF 27-JUL-1999; 99WO-GB02451.
 XX
 PR 27-JUL-1998; 98GB-0016337.
 PR 19-MAR-1999; 99US-0125164.
 XX
 PA (MICR-) MICROBIAL TECHNIQS LTD.
 PA
 PI Gilbert CFG, Hansbro PM;
 XX
 DR WPI: 2000-195300/17.
 XX
 PT New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing
 PT or inhibiting expression of the protein
 XX
 PS Claim 1; Page 77; 108pp; English.
 XX
 CC AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from Streptococcus pneumoniae. AAY05407 to AAY05590 represent
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
 CC The sequences have antibacterial and antiinflammatory properties.
 CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonising, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. AAY05591 to AAY05614 represent primers used in the
 CC exemplification of the present invention.

SQ Sequence 267 AA:

Query Match 73.1%; Score 38; DB 21; Length 267;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10
1:||||:11
DB 146 gatvdnldy 154

RESULT 6

ABB64149
ID ABB64149 standard; Protein; 345 AA.

XX ABB64149;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 19239.

XX Drosophila: developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL08252.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Disclosure: SEQ ID NO 19239; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA

XX sequences (AB101840-AB116175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 345 AA:

Query Match 73.1%; Score 38; DB 22; Length 345;
Best Local Similarity 70.0%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
1:||||:11
DB 31 essvvdslidy 40

RESULT 7

AAG65785
ID AAG65785 standard; Protein; 1956 AA.

XX AAG65785;

XX 11-FEB-2002 (first entry)

XX Human SNS1 ion channel protein.

XX Ion channel protein; gating; sensory neurone specific protein; SNS;

XX analgesic; antiinflammatory; immunosuppressive; human.

XX Homo sapiens.

XX WO200168681-A2.

XX 20-SEP-2001.

XX 14-MAR-2001; 2001WO-GB01108.

XX 14-MAR-2000; 2000GB-0006024.

XX (GLAX) GLAXO GROUP LTD.

XX Aitken AJ, Chen MX, Gladwell ZM, Powell AJ, Tate SN;

XX WPI; 2001-590038/66.

XX N-PSDB; AA16961.

XX Altered ion channel protein useful for screening modulators that are

XX useful in the manufacture of a medicament for the treatment of pain,

XX inflammation and hypersensitivity, has acquired sensitivity to gating

XX agent -

XX Disclosure: Page 57-58; 64pp; English.

XX The invention provides an altered ion channel protein having acquired

XX sensitivity to a gating agent. The altered ion channel protein can be

XX expressed by standard recombinant methodology. The altered ion channel

XX protein can be derived from a rat, mouse or human unaltered, gating agent

XX insensitive ion channel proteins. The altered protein is useful in

XX screening assay for the identification of modulators capable of

XX modulating an unaltered gating agent insensitive ion channel protein. The

XX modulator identified by the assay is preferably an antagonist of the

XX altered protein, which is useful in the manufacture of a medicament for

XX the treatment of pain, inflammation or hypersensitivity, and for treating

XX a mammalian patient, particularly human. The present sequence represents

XX a human sensory neurone specific (SNS) 1, a sodium ion channel protein.

SQ Sequence 1956 AA:

Query Match 71.2%; Score 37; DB 22; Length 1956;
Best Local Similarity 77.8%; Pred. No. 2,6e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMD 9
1:||||:1
DB 1085 egstvdsdld 1093

RESULT 8

AAB61996
ID AAB61996 standard; Protein; 1956 AA.

XX AAB61996;

XX 14-MAY-2001 (first entry)

XX Human peripheral nerve sodium channel type 3 (hPN3) protein.

XX Peripheral nerve tetrodotoxin-resistant sodium channel alpha subunit;

KM tetrodotoxin; TTX; peripheral nerve sodium channel type 3; PN3; human;
 KM sodium channel protein; peripheral nervous system; allodynia; neuropathy;
 KM hyperalgesia; traumatic injury; antidiabetic; neuroprotective; vulnerary;
 KM analgesic; anti-HIV.
 XX
 OS Homo sapiens.
 XX
 PN US6184349-B1.
 XX
 PD 06-FEB-2001.
 XX
 PF 15-APR-1997; 97US-0843417.
 XX
 PR 11-OCT-1995; 95US-0511828.
 XX
 PA (SYNTE) SYNTEX USA INC.
 XX
 PI Herman R, Delgado SG, Fish LM, Sangameswaran L, Rabert DK;
 XX
 DR WPI; 2001-202004/20.
 DR N-PSDB; AAF57016.
 XX
 PT New rat and human tetrodotoxin-resistant, voltage-gated sodium channel
 PT proteins, present in peripheral nerve tissue, useful as a therapeutic
 PT target for compounds treating peripheral nervous system disorders -
 XX
 PS Claim 2; Flg 6A-K; 86pp; English.
 XX
 CC The invention provides purified and isolated rat and human peripheral
 CC nerve tetrodotoxin (TTX)-resistant sodium channel alpha subunit proteins.
 CC The rat peripheral nerve sodium channel type 3 (PN3) protein or its human
 CC homologue are useful for the development of antibodies against PN3 which
 CC are useful in affinity chromatography to purify recombinant sodium
 CC channel proteins or polypeptides, or as a research tool. The PN3 proteins
 CC are useful as therapeutic targets for compounds to treat disorders of the
 CC peripheral nervous system such as allodynia, hyperalgesia, diabetic
 CC neuropathy, traumatic injury and acquired immunodeficiency syndrome
 CC (AIDS)-associated neuropathy. The present sequence represents the human
 CC hPN3 protein.
 CC
 SQ Sequence 1956 AA;
 XX
 XX
 Query Match 71.2%; Score 37; DB 22; Length 1956;
 Best Local Similarity 77.8%; Pred. No. 2.6e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 OY 1 EGSTVDSMD 9
 Db 1085 egstvdcl d 1093
 XX
 RESULT 9
 ID ABB59854 standard; Protein; 2005 AA.
 XX
 AC ABB59854;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 6354.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX

PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL03957.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 6354; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2005 AA;
 XX
 XX
 Query Match 69.2%; Score 36; DB 22; Length 2005;
 Best Local Similarity 66.7%; Pred. No. 4.3e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 OY 2 GSTVDSMDY 10
 Db 11 gngidsmdy 19
 XX
 RESULT 10
 AAU71866
 ID AAU71866 standard; Protein; 275 AA.
 XX
 AC AAU71866;
 XX
 DT 26-FEB-2002 (first entry)
 XX
 DE C. glutamicum metabolic pathway protein encoded by gene #1.
 XX
 KM Metabolic pathway protein; MP; lysine biosynthesis pathway;
 KM methionine biosynthesis pathway; large-scale production of fine chemical;
 KM Corynebacterium diptheriae; diptheria.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200166573-A2.
 XX
 PD 13-SEP-2001.
 XX
 PF 22-DEC-2000; 2000WO-IB02035.
 XX
 PR 09-MAR-2000; 2000US-187970P.
 PR 23-JUN-2000; 2000US-0606740.
 XX
 PA (BADT) BASF AG.
 XX
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;
 PI Lee H, Hwang B;
 XX
 DR WPI; 2001-582269/65.
 DR N-PSDB; AAS96076.
 XX


```

RESULT 12
AAC91878
ID AAG91878 standard; Protein; 277 AA.
XX
AC AAG91878;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 5632.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI: 2001-376931/40.
DR N-PSDB: AAH67097.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX
PS Claim 17: SEQ ID NO: 5632; 246pp + Sequence Listing; English.
XX
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX
SQ Sequence 277 AA;

```

```

Query Match          67.3%; Score 35; DB 22; Length 277;
Best Local Similarity 64.3%; Pred. NO. 67;
Matches 9; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

```

```

OY 1 EGSTVDS---MDY 10
   |||||
   |||||
Db 59 egstvdpslwfmdy 72

```

```

RESULT 13
AAB70804
ID AAB70804 standard; Protein; 277 AA.
XX
AC AAB70804;
XX

```

```

DT 04-JUN-2001 (first entry)
XX
XX C. glutamicum dapF protein.
DE
XX Overexpression; fermentation; L-lysine; lysine biosynthesis;
KW diaminopimelate epimerase; dapF.
XX
XX Corynebacterium glutamicum.
XX
XX DE1943587-A1.
XX
XX 15-MAR-2001.
XX
XX 11-SEP-1999; 99DE-1043587.
XX
XX 11-SEP-1999; 99DE-1043587.
XX
XX (DEGS ) DEGUSSA-HUELS AG.
XX
XX Kirchner O, Bathe B, Moeckel B, Hartmann M, Kalinowski J;
PI Puehler A, Pfefferle W;
XX
XX WPI: 2001-258936/27.
XX
XX N-PSDB: AAF61391.
XX
XX
XX Cloned Corynebacterium glutamicum diaminopimelate epimerase (dapF) gene
XX useful for producing dapF-overexpressing coryneform bacteria for the
XX fermentative production of L-lysine
XX
XX
XX Claim 6; Page 13; 18pp; German.
XX
XX
XX This invention describes a novel cloned Corynebacterium glutamicum
XX diaminopimelate epimerase (dapF), EC 5.1.1.7 (I). The invention also
XX describes (a) a polynucleotide that is at least 70% identical to a
XX polynucleotide encoding a polypeptide with a defined sequence of 277
XX amino acids (II) given in the specification; (b) a polynucleotide that is
XX at least 70% identical to a polynucleotide encoding a polypeptide
XX expressed by the dapF gene in pEC-X799A-dapF in C. glutamicum DSM 12968;
XX (c) a polynucleotide encoding a polypeptide comprising an amino acid
XX sequence that is at least 70% identical to (II); (d) a polynucleotide
XX complementary to the polynucleotide of (a), (b) or (c); and (e) a
XX polynucleotide comprising at least 15 consecutive bases of the
XX polynucleotide sequence of (a), (b), (c) or (d). (I) can be used to
XX produce dapF-overexpressing coryneform bacteria for the fermentative
XX production of L-lysine.
XX
XX
XX Sequence 277 AA;

```

```

Query Match          67.3%; Score 35; DB 22; Length 277;
Best Local Similarity 64.3%; Pred. NO. 67;
Matches 9; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

```

```

OY 1 EGSTVDS---MDY 10
   |||||
   |||||
Db 59 egstvdpslwfmdy 72

```

```

RESULT 14
AAM93294
ID AAM93294 standard; Protein; 371 AA.
XX
AC AAM93294;
XX

```

```

DT 06-NOV-2001 (first entry)
XX

```

```

DE Human polypeptide, SEQ ID NO: 2787.
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
XX Homo sapiens.
XX
XX EPI130094-A2.
XX

```

XX 05-SEP-2001.
PD
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
XX 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
DR N-PSDB; AAK94214.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2787; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
CC
XX
SQ Sequence 371 AA:

Query Match 67.3%; Score 35; DB 22; Length 371;
Best Local Similarity 60.0%; Pred. No. 94;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
||:|:|
Db 87 egatldity 96

RESULT 15
ABG26659
ID ABG26659 standard; Protein: 473 AA.
XX
AC ABG26659;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #26650.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS90846.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 57018; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
CC
XX
SQ Sequence 473 AA:

Query Match 67.3%; Score 35; DB 22; Length 473;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
||:|:|
Db 145 egsvirslay 154

Search completed: August 20, 2002, 13:15:11
Job time: 105 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:16:02 ; Search time 46.28 Seconds
(without alignments)
20.763 Million cell updates/sec

Title: US-09-824-286-3_COPY_95_104
Perfect score: 52
Sequence: 1 EGSVDSMDY 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	75.0	1078	2 PC4198	peptide synthetase
2	38	73.1	262	2 E95164	prolipo protein dia
3	38	73.1	266	2 D98030	prolipo protein dia
4	37	71.2	329	2 A49868	cathepsin K (EC 3.
5	35	67.3	260	2 T09023	hypothetical prote
6	35	67.3	280	2 T15813	hypothetical prote
7	35	67.3	351	2 C91182	probable fibrillar
8	35	67.3	351	2 G86028	probable fibrillar
9	35	67.3	484	2 T33492	hypothetical prote
10	34	65.4	257	2 D82119	hypothetical prote
11	34	65.4	268	2 F64024	hypothetical prote
12	34	65.4	274	2 F72689	hypothetical prote
13	34	65.4	617	2 AF2897	conserved hypothet
14	34	65.4	641	2 H97672	hypothetical prote
15	34	65.4	822	2 S56823	probable membrane
16	34	65.4	1210	2 H88451	protein ZC135.3 (I
17	33.5	64.4	303	2 B95101	1-phosphofructokin
18	33.5	64.4	303	2 C97969	1-phosphofructokin
19	33	63.5	74	2 A30475	hypothetical prote
20	33	63.5	99	1 OPHUE	acylphosphatase (E
21	33	63.5	134	2 J01403	acidic seminal flu
22	33	63.5	151	1 GGGACR	globin - horn shel
23	33	63.5	162	2 A87755	protein T21E12.3 (
24	33	63.5	245	2 T03088	conserved hypothet
25	33	63.5	269	2 S61555	xyloglucan endo-1,
26	33	63.5	364	2 E83995	leucine dehydrogen
27	33	63.5	400	2 S29686	penicillin-binding
28	33	63.5	400	2 S29687	penicillin-binding
29	33	63.5	400	2 S49089	penicillin-binding

30	33	63.5	401	2 S29688	penicillin-binding
31	33	63.5	518	2 T24783	hypothetical prote
32	33	63.5	536	2 D84325	Htt17 transducer (
33	33	63.5	680	2 T01403	hypothetical prote
34	33	63.5	682	2 U00420	beta-1,3-glucanase
35	33	63.5	694	2 H95012	hypothetical prote
36	33	63.5	702	2 C97884	hypothetical prote
37	33	63.5	785	2 H85036	hypothetical prote
38	33	63.5	1025	2 S34839	collagen alpha 1(V
39	33	63.5	1957	2 S68453	sodium channel pro
40	32	61.5	100	1 A61449	acylphosphatase (E
41	32	61.5	111	2 T37114	hypothetical prote
42	32	61.5	145	2 S75719	molycoprotein bios
43	32	61.5	187	2 A41489	hypothetical prote
44	32	61.5	246	2 A43579	Vmp7 protein homol
45	32	61.5	260	2 E89965	enterotoxin Seo (I

ALIGNMENTS

```

RESULT 1
PC4198
Peptide synthetase (EC 1.14.17.-) - Imperfect fungus (Metarhizium anisopliae) (fragme
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 03-Nov-2000
C:Accession: PC4198; PC4199
R:Bailey, A.M.; Kershaw, M.J.; Hunt, B.A.; Paterson, I.C.; Charnley, A.K.; Reynolds,
Gene 173, 195-197, 1996
A:Title: Cloning and sequence analysis of an intron-containing domain from a peptide
A:Reference number: PC4198; MUID:97082966
A:Accession: PC4198
A:Molecule type: DNA
A:Residues: 1-1078 <BAI>
A:Cross-references: EMBL:X89442; NID:q1015536; PID:q1015537
A:Accession: PC4199
A:Molecule type: Protein
A:Residues: 58-67;136-146 <BI2>
C:Genetics:
A:Gene: pes
A:Introns: 386/3; 762/2
C:Superfamily: acyl carrier protein homology; acetate--CoA ligase homology
C:Keywords: carrier protein; oxidoreductase
F:32-499/Domain: acetate--CoA ligase homology <ACI>
F:526-594/Domain: acyl carrier protein homology <ACP3>

Query Match          75.0%; Score 39; DB 2; Length 1078;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGSVDSMD 9
DB 438 DGSVDSLD 446

RESULT 2
E95164
Prolipo protein diacylglycerol transferase [imported] - Streptococcus pneumoniae (stra
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: E95164
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umeyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: E95164
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-262 <KUR>

```

A:Cross-references: GB:AE005672; PIDN:AAK75510.1; PID:q14972901; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPL412
C:Superfamily: prolipoprotein diacylglycerol transferase

Query Match 73.1%; Score 38; DB 2; Length 262;
Best Local Similarity 66.7%; Pred. No. 7.7;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 2 GSTVDSMDY 10
DB 142 GATVDNLDY 150

RESULT 3
D98030
prolipoprotein diacylglycerol transferase (EC 2.4.99.-) [imported] - Streptococcus pneum
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D98030
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; F
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A57872; MUID:21429245; PMID:11544234
A:Accession: D98030
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-266 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAL00073.1; PID:g15458909; GSPDB:GN00174
C:Genetics:
A:Gene: lgt
C:Superfamily: prolipoprotein diacylglycerol transferase
C:Keywords: glycosyltransferase

Query Match 73.1%; Score 38; DB 2; Length 266;
Best Local Similarity 66.7%; Pred. No. 7.8;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 2 GSTVDSMDY 10
DB 146 GATVDNLDY 154

RESULT 4
AA9868
cathepsin K (EC 3.4.22.-) precursor [similarity] - rabbit
N:Alternate names: osteoclast cysteine proteinase OC-2
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 30-Jun-1995 #sequence_revision 30-Jun-1995 #text_change 20-Jun-2000
C:Accession: AA9868
R:Tezuka, K.; Tezuka, Y.; Maejima, A.; Sato, T.; Nemoto, K.; Kamiooka, H.; Hakeda, Y.; Ku
J. Biol. Chem. 269, 1106-1109, 1994
A>Title: Molecular cloning of a possible cysteine proteinase predominantly expressed in
A:Reference number: AA9868; MUID:94117413
A:Accession: AA9868
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-329 <TEZ>
A:Cross-references: GB:D14036; NID:g454186; PIDN:BA003125.1; PID:g454187
C:Superfamily: papain
C:Keywords: cysteine proteinase; hydrolase
E:139,276,296/Active site: Cys, His, Asn #status predicted

Query Match 71.2%; Score 37; DB 2; Length 329;
Best Local Similarity 70.0%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
DB 112 EGRTDPSIDY 121

RESULT 5
T09023
hypothetical protein T27E11.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C:Accession: T09023
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ban
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T09023
A:Molecule type: DNA
A:Residues: 1-260 <BEV>
A:Cross-references: EMBL:AL078579; GSPDB:GN00062; ATSP:T27E11.80
A:Experimental source: cultivar Columbia; BAC clone T27E11
C:Genetics:
A:Gene: ATSP:T27E11.80
A:Map position: 4

Query Match 67.3%; Score 35; DB 2; Length 260;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGSTVDSMD 9
DB 98 DGTADSLD 106

RESULT 6
T15813
hypothetical protein C46H3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15813
R:Favell, T.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid C46H3.
A:Reference number: Z16409
A:Accession: T15813
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-280 <FAV>
A:Cross-references: EMBL:U41271; NID:g1086855; PID:g1086858; PIDN:AAA82445.1; CESP:C4
C:Genetics:
A:Gene: CESP:C46H3.3
A:introns: 7/3; 59/1; 107/1; 155/3; 187/3; 224/3; 263/3

Query Match 67.3%; Score 35; DB 2; Length 280;
Best Local Similarity 66.7%; Pred. No. 33;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 GSTVDSMDY 10
DB 130 GANVDSVDY 138

RESULT 7
C91182
probable fibrinolytic protein precursor [imported] - Escherichia coli (strain 0157:H7, su
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C91182
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C91182
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-351 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA837850.1; PID:913363901; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs4427

Query Match 67.3%; Score 35; DB 2; Length 351;
Best Local Similarity 77.8%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10
| | | | | | |
Db 99 GRTVDSMKY 107

RESULT 8
G86028
probable fimbrial protein 24966 [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G86028

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G86028
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-351 <STO>
A:Cross-references: GB:AE005174; NID:912518274; PIDN:AAG58691.1; GSPDB:GN00145; UWGP:Z49
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 24966

Query Match 67.3%; Score 35; DB 2; Length 351;
Best Local Similarity 77.8%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10
| | | | | | |
Db 99 GRTVDSMKY 107

RESULT 9
T33492
hypothetical protein C36C9.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
C:Accession: T33492

R:Fullen, B.; Martin, J.; O'Brien, D.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid C36C9.
A:Reference number: Z21357

A:Accession: T33492
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-484 <FNU>
A:Cross-references: EMBL:AF098986; PIDN:AAC67427.1; GSPDB:GN00028; CESP:C36C9.5
A:Experimental source: strain Bristol N2; clone C36C9

C:Genetics:
A:Gene: CESP:C36C9.5
A:Map position: X
A:Introns: 31/1; 136/1; 194/1; 257/1; 282/1; 313/1; 344/1; 373/1; 408/1; 437/3
C:Superfamily: Caenorhabditis elegans hypothetical protein C36C9.5

Query Match 67.3%; Score 35; DB 2; Length 484;
Best Local Similarity 60.0%; Pred. No. 58;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
| | | | | | |
Db 235 EGSSSETMDY 244

RESULT 10
D82119
hypothetical protein VC2094 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82119

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833
A:Accession: D82119
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-257 <HEI>
A:Cross-references: GB:AE004283; GB:AE003852; NID:99656636; PIDN:AAF95240.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Map position: 1

Query Match 65.4%; Score 34; DB 2; Length 257;
Best Local Similarity 60.0%; Pred. No. 47;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
| | | | | | |
Db 84 EGKTVGSIDF 93

RESULT 11
F64024
hypothetical protein H1273 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: F64024

R:Flatschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weisman
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:9350630
A:Accession: F64024

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-268 <TIGR>
A:Cross-references: GB:U32807; GB:LA42023; NID:91574725; PIDN:AAC22921.1; PID:91574728

Query Match 65.4%; Score 34; DB 2; Length 268;
Best Local Similarity 60.0%; Pred. No. 49;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
| | | | | | |
Db 80 OGSTVYALDY 89

RESULT 12
F72689
hypothetical protein APE0934 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C:Accession: F72689
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahara, H.; Takamaya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339
A:Accession: F72689
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-274 <KAN>
A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79918.1; PID:d1043704; PID:g5104188
A:Experimental source: strain KI
C:Genetics:
A:Gene: APE0934
C:Superfamily: Aeropyrum pernix hypothetical protein APE0934

Query Match 65.4%; Score 34; DB 2; Length 274;
Best Local Similarity 60.0%; Pred. No. 50;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
|||||:
Db 241 EGSTDSIPF 250

RESULT 13
AF2897
Conserved hypothetical protein Atu2615 [Imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AF2897
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, S.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, A.; Sterner, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF2897
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-617 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL43596.1; PID:g17741114; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu2615
A:Map position: circular chromosome

Query Match 65.4%; Score 34; DB 2; Length 617;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
|||||:
Db 336 EASSVEGMDY 345

RESULT 14
H97672
hypothetical protein AGR_C_4742 [Imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: H97672
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194

A:Accession: H97672
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-641 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK88337.1; PID:g15157817; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_4742
A:Map position: circular chromosome

Query Match 65.4%; Score 34; DB 2; Length 641;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
|||||:
Db 360 EASSVEGMDY 369

RESULT 15
S56823
probable membrane protein YJL051w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein J1156
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 09-Jun-2000
C:Accession: S56823
R:Pohl, T.M.; Aljinovic, G.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56793
A:Accession: S56823
A:Molecule type: DNA
A:Residues: 1-822 <TOV>
A:Cross-references: EMBL:Z49326; NID:g1008186; PID:g1008187; GSPDB:GN00010; MIPS:YJL051w
C:Genetics:
A:Gene: MIPS:YJL051w
A:Map position: 10L
C:Superfamily: Saccharomyces cerevisiae probable membrane protein YJL051w
C:Keywords: transmembrane protein

Query Match 65.4%; Score 34; DB 2; Length 822;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
|||||:
Db 670 EGSTDPINW 679

Search completed: August 20, 2002, 13:16:03
Job time: 157 sec

1000 2000 3000 4000 5000 6000 7000 8000 9000 10000

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:51 ; Search time 21.48 Seconds

(without alignments)
18.026 Million cell updates/sec

Title: US-09-824-286-3_COPY_95_104
Perfect score: 52
Sequence: 1 EGSTVDSMDY 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	71.2	329	1	CATK_RABIT
2	37	71.2	330	1	CATK_PIG
3	34	65.4	268	1	YC73_HAEIN
4	34	65.4	822	1	YJF1_YEAST
5	33	63.5	98	1	ACYO_MOUSE
6	33	63.5	134	1	ASFP_BOVIN
7	33	63.5	151	1	GAB_CERRH
8	33	63.5	247	1	MER5_ARATH
9	33	63.5	366	1	DHLE_BACE
10	33	63.5	421	1	ECB2_HAEL
11	33	63.5	518	1	YRVA_CAEEL
12	33	63.5	682	1	E13B_BACCT
13	33	63.5	836	1	PDC2_CANAL
14	33	63.5	1025	1	CA16_MOUSE
15	32	61.5	98	1	ACYO_MOUSE
16	32	61.5	100	1	ACYO_BOVIN
17	32	61.5	100	1	ACYO_PIG
18	32	61.5	145	1	Y903_STYR3
19	32	61.5	329	1	CATK_HUMAN
20	32	61.5	329	1	CATK_MACRA
21	32	61.5	329	1	CATK_MOUSE
22	32	61.5	329	1	CATK_RAT
23	32	61.5	587	1	Y548_CAEEL
24	32	61.5	628	1	ABRA_ASPMG
25	32	61.5	682	1	NISP_YEAST
26	32	61.5	751	1	YK09_YEAST
27	31	59.6	160	1	GREM_MYCPN
28	31	59.6	161	1	GREM_MYCPN
29	31	59.6	207	1	JAG_BACCH
30	31	59.6	229	1	YXK8_CAEEL
31	31	59.6	289	1	YFCB_YEAST
32	31	59.6	316	1	MCH_ARCFU
33	31	59.6	363	1	YR62_CAEEL

34	31	59.6	397	1	TLG2_YEAST	008144 saccharomyc
35	31	59.6	417	1	GAB3_YEAST	004739 saccharomyc
36	31	59.6	428	1	LE22_METTH	027668 methanobact
37	31	59.6	458	1	YC99_YEAST	P25594 saccharomyc
38	31	59.6	472	1	C15Y_DAUCA	080433 daucus caro
39	31	59.6	486	1	MORE_BACCH	09K954 bacillus ha
40	31	59.6	509	1	HMD1_ORYSA	P48019 oryza sativ
41	31	59.6	523	1	SYN_TREPA	083618 treponema p
42	31	59.6	582	1	YK85_YEAST	P36172 saccharomyc
43	31	59.6	769	1	ITB8_HUMAN	P26012 homo sapien
44	31	59.6	878	1	SYV_METJA	058413 methanococc
45	31	59.6	966	1	ST10_MOUSE	055098 mus musculu

ALIGNMENTS

RESULT	ID	CATK_RABIT	STANDARD	PRT	329 AA.
AC	P43236				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Cathepsin K precursor (EC 3.4.22.38) (OC-2 protein).				
GN	CTSK.				
OS	Oryctolagus cuniculus (Rabbit).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.				
OX	NCBI_TaxID=9986;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Bone;				
RX	MEDLINE=94117413; PubMed=8288568;				
RA	Tezuka K.-I., Tezuka Y., Maejima A., Sato T., Nemoto K.,				
RA	Kamiooka H., Hakeda Y., Kamegawa M.;				
RT	"Molecular cloning of a possible cysteine proteinase predominantly expressed in osteoclasts.";				
RL	J. Biol. Chem. 269:1106-1109(1994).				
CC	- FUNCTION: CLOSELY INVOLVED IN OSTEOCLASTIC BONE RESORPTION AND MAY PARTICIPATE PARTIALLY IN THE DISORDER OF BONE REMODELING. DISPLAYS POTENT ENDOPEPTIDASE ACTIVITY AGAINST FIBRINOGEN AT ACID PH. MAY PLAY AN IMPORTANT ROLE IN EXTRACELLULAR MATRIX DEGRADATION.				
CC	- CATALYTIC ACTIVITY: Broad proteolytic activity. With small-molecule substrates and inhibitors, the major determinant of specificity is P2, which is preferably Leu, Met > Phe, and not Arg.				
CC	- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN OSTEOCLASTS (BONES).				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE PAPAEN FAMILY OF THIOLE PROTEASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL: D14036; BAA03125.1; -				
DR	HSSP: P43235; IMEM.				
DR	MEROPS: C01.036; -				
DR	InterPro: IPR000668; Peptidase.C1.				
DR	InterPro: IPR000169; Thiolprol_act_site.				
DR	Pfam: PF00112; Peptidase.C1; 1.				
DR	PRINTS: PRO0705; PAPAEN.				
DR	PROSITE: PS00139; THIOLE_PROTEASE_CYS; 1.				
DR	PROSITE: PS00639; THIOLE_PROTEASE_HIS; 1.				
DR	PROSITE: PS00640; THIOLE_PROTEASE_ASN; 1.				
KW	Hydrolase; Thiol protease; Zymogen; Signal.				
KW	SIGNAL				
FT	PROPEP	16	114		ACTIVATION PEPTIDE.
FT	CHAIN	115	329		CATHEPSIN K.

```

FT ACT_SITE 139 139 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT ACT_SITE 296 296 BY SIMILARITY.
FT DISULFID 136 177 BY SIMILARITY.
FT DISULFID 170 210 BY SIMILARITY.
FT DISULFID 269 318 BY SIMILARITY.
FT CAROHND 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHND 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 329 AA: 36870 MW: 875D8582876B51F CRC64:

Query Match
Best Local Similarity 71.2%; Score 37; DB 1; Length 329;
Best Local Similarity 70.0%; Pred. No. 4.8;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
DB 112 EGRTPDSIDY 121

RESULT 2
CATK_PIG STANDARD; PRT: 330 AA.
AC 09GLE3:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cathepsin K precursor (EC 3.4.22.38).
GN CTSK.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RC TISSUE=Thyroid;
RX MEDLINE=20534869; PubMed=11082042;
RA Tepel C., Broemme D., Herzog V., Brix K.;
RA "Cathepsin K in thyroid epithelial cells: sequence, localization and
RT possible function in extracellular proteolysis of thyroglobulin.";
RL J. Cell Sci. 113:4487-4498(2000).
CC -1- FUNCTION: Closely involved in the disorder of bone remodeling and may
CC participate partially in the disorder of bone remodeling. Displays
CC potent endoprotease activity against fibrinogen at acid pH. May
CC play an important role in extracellular matrix degradation (by
CC similarity). Mediates thyroxine liberation by limited proteolysis
CC of thyroglobulin at neutral pH in vitro.
CC -1- CATALYTIC ACTIVITY: Broad proteolytic activity. With small-
CC molecule substrates and inhibitors, the major determinant of
CC specificity is P2, which is preferably Leu, Met > Phe, and not
CC Arg.
CC -1- SUBCELLULAR LOCATION: Lysosomal and extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC PAPAIN FAMILY OF THIOL PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, AF292030; AAC12340.1; -.
CC InterPro: IPR000668; Peptidase_C1.
CC InterPro: IPR000169; Thiolprol_act_site.
CC Pfam: PF00112; Peptidase_C1; 1.
CC PRINTS: PR00705; PAPAIN
CC PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
CC PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
CC PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.
CC Hydrolase: Thiol protease; Lysosome; Signal; Lysosome; Glycoprotein.
FT SIGNAL 1 16 POTENTIAL.

```

```

FT PROPEP 17 115 ACTIVATION PEPTIDE.
FT CHAIN 116 330 CATHEPSIN K.
FT ACT_SITE 140 140 BY SIMILARITY.
FT ACT_SITE 277 277 BY SIMILARITY.
FT ACT_SITE 297 297 BY SIMILARITY.
FT DISULFID 137 178 BY SIMILARITY.
FT DISULFID 171 211 BY SIMILARITY.
FT DISULFID 270 319 BY SIMILARITY.
FT CAROHND 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 330 AA: 37069 MW: ABE089EFE5FB170 CRC64:

Query Match
Best Local Similarity 71.2%; Score 37; DB 1; Length 330;
Best Local Similarity 70.0%; Pred. No. 4.8;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
DB 113 EGRTPDSIDY 122

RESULT 3
YC73_HAEIN STANDARD; PRT: 268 AA.
AC P44150;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein H11273.
GN H11273
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Keriavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Georgagen N.S.M.,
RA Gnehm C.L., McDonald J.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, U32807; AAC22921.1; -.
CC TIGR: H11273;
CC InterPro: IPR000051; SAM_bind.
CC Hypothetical protein; Complete proteome.
KW SEQUENCE 268 AA: 30510 MW: E5B28DA7AAD4D0B CRC64:

Query Match
Best Local Similarity 65.4%; Score 34; DB 1; Length 268;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
DB 80 QGSTVVALDY 89

```

```

RESULT 4
ID YJFL_YEAST STANDARD: PRT: 822 AA.
AC P47046.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 91.3 kDa protein in TDPH-MTR4 intergenic region.
YJL051W OR J1156.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pohl T.M., Aljinovic G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49326; CAAB9342.1; -.
DR SGD; S0003587; YJL051W.
DR Hypothetical protein; Transmembrane.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 77 97 POTENTIAL.
FT TRANSMEM 121 141 POTENTIAL.
SQ SEQUENCE 822 AA; 91256 MW; BB850C5B409F8820 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 822;
Best Local Similarity 60.0%; Pred. NO. 55;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSYVDSMDY 10
Db 670 EGSTDPINTY 679

RESULT 5
ID ACYO_HUMAN STANDARD: PRT: 98 AA.
AC P07311.
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acylphosphatase, organ-common type isozyme (EC 3.6.1.7)
DE (Acylphosphate phosphohydrolase) (Acylphosphatase, erythrocyte
DE isozyme)
GN ACPI OR ACYPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RA MEDLINE=87101109; PubMed=3026468;
RA Liguri G., Camici G., Manzo G., Cappugi G., Nassi P., Modesti A.,
RA Ramponi G.;
RT "A new acylphosphatase isoenzyme from human erythrocytes:
RT purification, characterization, and primary structure.";
RL Biochemistry 25:8089-8094(1986).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Placenta;
RA Raugel G.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Dors M., Dickhoff R., James R.,
RA Loretz C., Laaky S., Madan A., Prescott S., Ratcliffe A., Shaffer T.,
RA Hood L.;
RT "Sequencing of human chromosome 14q24.3 region.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 9-85 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95317414; PubMed=7796909;
RA Flaschi T., Raugel G., Marzocchi R., Chiarugi P., Cirri P.,
RA Ramponi G.;
RT "Cloning and expression of the cDNA coding for the erythrocyte
RT isoenzyme of human acylphosphatase.";
RL FEBS Lett. 367:145-148(1995).
CC -1- FUNCTION: ITS PHYSIOLOGICAL ROLE IS NOT YET CLEAR.
CC -1- CATALYTIC ACTIVITY: An acyl phosphate + H(2)O = a fatty acid anion
CC + phosphate
CC -1- TISSUE SPECIFICITY: ORGAN-COMMON TYPE ISOZYME IS FOUND IN MANY
CC DIFFERENT TISSUES.
CC -1- SIMILARITY: BELONGS TO THE ACYLPHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X84194; CA58987.1; -.
DR EMBL; AC007055; AAD31937.1; -.
DR PIR; A25587; OPHOE.
DR HSSP; P41500; ZACY.
DR MIM; 600875; -.
DR InterPro: IPR001792; Acylphosphatase.
DR Pfam: PF00708; Acylphosphatase; 1.
DR PRINTS; PR00112; ACYLPHPTASE.
DR ProDom; PD001884; Acylphosphatase; 1.
DR PROSITE; PS00150; ACYLPHOSPHATASE_1; 1.
DR PROSITE; PS00151; ACYLPHOSPHATASE_2; 1.
KM Hydrolyase; Acetylation; Multigene family.
FT INIT_MET 0 0
FT MOD_RES 1 1 ACETYLATION.
FT ACT_SITE 23 23 POTENTIAL.
FT ACT_SITE 41 41 POTENTIAL.
FT CONFLICT 19 19 G -> R (IN REF. 4)
SQ SEQUENCE 98 AA; 11130 MW; 6C5EE71D00258B06 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 98;
Best Local Similarity 60.0%; Pred. NO. 8;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGSYVDSMDY 10
Db 2 EGNLTISVDY 11

RESULT 6
ID ASFP_BOVIN STANDARD: PRT: 134 AA.
AC P29392.
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Acidic seminal fluid protein precursor (ASFP).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 OK NCBI_TaxID=9913;
 RN (1)
 RN SEQUENCE FROM N.A.
 RC TISSUE=Seminal vesicle;
 RX MEDLINE=92181448; PubMed=1543494;
 RA Wempe F., Einspanier R., Scheit K.H.;
 RT "Characterization by cDNA cloning of the mRNA of a new growth factor
 from bovine seminal plasma: acidic seminal fluid protein.";
 RL Biochem. Biophys. Res. Commun. 183:232-237(1992).
 RN (2)
 RP SEQUENCE OF 21-43.
 RC TISSUE=Seminal vesicle;
 RX MEDLINE=91378963; PubMed=1898381;
 RA Einspanier R., Einspanier A., Wempe F., Scheit K.H.;
 RT "Characterization of a new bioactive protein from bovine seminal
 fluid.";
 RL Biochem. Biophys. Res. Commun. 179:1006-1010(1991).
 RN (3)
 RP DISULFIDE BONDS.
 RX MEDLINE=94237283; PubMed=8181566;
 RA Einspanier R., Krause I., Calvete J.J., Toepfer-Petersen E.,
 RT Klostermeyer H., Karg H.;
 RT "Bovine seminal plasma asFP: localization of disulfide bridges and
 detection of three different isoelectric forms.";
 RL FEBS Lett. 344:61-64(1994).
 RN (4)
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RX MEDLINE=97475216; PubMed=9334740;
 RA Romero A., Romao M.J., Varela P.F., Koelln I., Dias J.M.,
 RA Carvalho A.L., Sanz L., Toepfer-Petersen E., Calvete J.J.;
 RT "The crystal structures of two spermadhesins reveal the CUB domain
 fold.";
 RL Nat. Struct. Biol. 4:783-788(1997).
 CC -1- FUNCTION: STIMULATES CELL DIVISION AND PROGESTERONE SECRETION
 OF BOVINE GRANULOSA CELLS IN VITRO IN A POTENT AND DOSE DEPENDENT
 MANNER. THIS PROTEIN APPEARS TO BE A POTENT GROWTH FACTOR WITH
 EFFECTS ON OVARIAN GRANULOSA CELLS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: SEMINAL VESICLE TISSUE, AMPULLA AND WEAKLY
 IN TISSUE OF EPIDIDYMIS.
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SPERMADHESIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL: M84603; AAA0745.1; .
 DR PIR: PH0213; PH0213.
 DR PIR: J01403; J01403.
 DR PDB: 1SFP; 24-JUN-98.
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR000124; Spermadhesin.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.
 DR PROSITE: PS00985; SPERMADHESIN_1; 1.
 DR PROSITE: PS00986; SPERMADHESIN_2; 1.
 DR PROSITE: PS01180; CUB; 1.
 DR Growth factor; Signal; 3D-structure.
 KW SIGNAL
 FT SIGNAL 1 20
 FT CHAIN 21 134 ACIDIC SEMINAL FLUID PROTEIN.
 FT DOMAIN 30 131 CUB.
 FT DISULFID 30 51
 FT DISULFID 74 95
 FT CONFLICT 43 43 T -> H (IN REF. 2).
 SQ SEQUENCE 134 AA; 15036 MW; 339BCFF8637D64C0 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 134;
 Best Local Similarity 75.0%; Pred. No. 11;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 STYVDSMDY 10
 Db 16 ATYVDSMDW 23
 RESULT 7
 GLB_CERRH STANDARD; PRT; 151 AA.
 AC P02215;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE Globin (Myoglobin).
 OS Cerithidea rhizophorarum (Water snail) (Horn shell).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
 OC Neoenioglossa; Cerithioidea; Potamididae; Cerithidea.
 OK NCBI_TaxID=6472;
 RN (1)
 RP SEQUENCE.
 RC TISSUE=Radular muscle;
 RX MEDLINE=83204919; PubMed=6849938;
 RA Takagi T., Tobita M., Shikama K.;
 RT "Amino acid sequence of dimeric myoglobin from Cerithidea
 rhizophorarum.";
 RL Biochim. Biophys. Acta 745:32-36(1983).
 CC -1- SUBUNIT: HOMODIMER.
 PIR: A02537; GGGACR.
 DR HSSP: P14821; 1SCF.
 DR InterPro: IPR002336; Erythcrutin.
 DR InterPro: IPR000971; Globin.
 DR Pfam: PF00042; globin; 1.
 DR PRINTS: PR00611; ERYTHCRUTIN.
 DR PROSITE: PS01033; GLOBIN; 1.
 KW Heme; Oxygen transport; Transport; Muscle; Acetylation.
 FT MOD_RES 1 1
 FT METAL 66 66 ACETYLATION.
 FT METAL 98 98 IRON (HEME DISTAL LIGAND)
 FT METAL 98 98 (BY SIMILARITY).
 FT METAL 98 98 IRON (HEME PROXIMAL LIGAND)
 FT METAL 98 98 (BY SIMILARITY).
 SQ SEQUENCE 151 AA; 16210 MW; 3493BAE8F44BD90 CRC64;
 Query Match 63.5%; Score 33; DB 1; Length 151;
 Best Local Similarity 75.0%; Pred. No. 13;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Oy 2 GSTVDSMD 9
 Db 75 GSWDSMD 82
 RESULT 8
 MERS_ARATH STANDARD; PRT; 247 AA.
 ID MERS_ARATH
 AC P24806;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-MAR-1992 (Rel. 21, Last annotation update)
 DE MERI-5 protein.
 GN MERI-5.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OK NCBI_TaxID=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93005704; PubMed=1840916;

RA Medford J.I., Elmer J.S., Klee H.J.;
 RT "Molecular cloning and characterization of genes expressed in shoot
 apical meristems."
 RL Plant Cell 3:359-370(1991).
 CC -I- TISSUE SPECIFICITY: EXPRESSED IN SHOOT APICAL MERISTEMS, ALSO
 CC FOUND IN SEEDLINGS AND MERISTEMS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M63166; AAA32828.1; -
 DR PIR: J01022; J01022.
 DR HSSP: P23904; IAKK.
 DR InterPro: IPR000757; Glyco_hydro_16.
 DR Pfam: PF00722; Glyco_hydro_16; 1.
 SQ SEQUENCE 247 AA; 28295 MW; 831E8441564B45E8 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 247;
 Best Local Similarity 60.0%; Pred. No. 23;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
 |||||:|:
 Db 92 EGSTWDEIDF 101

RESULT 9
 DHLE_BACCE STANDARD; PRT; 366 AA.
 ID DHLE_BACCE
 AC Q93194;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Leucine dehydrogenase (EC 1.4.1.9) (LeudH).
 GN LDH.
 OS Bacillus cereus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1396;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 626;
 RA Stoyan T.; Recktenwald A.; Kula M.R.;
 RT "Cloning, sequencing and overexpression of the leucine dehydrogenase
 RT gene from Bacillus cereus."
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: FUNCTIONS CATABOLICALLY IN THE BACTERIAL METABOLISM OF
 CC BRANCHED-CHAIN L-AMINO ACIDS, AND PLAYS AN IMPORTANT ROLE IN SPORE
 CC GERMINATION IN COOPERATION WITH ALANINE DEHYDROGENASE (BY
 CC SIMILARITY).
 CC -I- CATALYTIC ACTIVITY: L-leucine + H(2)O + NAD(+) = 4-methyl-2-
 CC oxopentanoate + NH(3) + NADH
 CC -I- SUBUNIT: HOMOHXAMER (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U51099; AAA96314.1; -
 DR HSSP: P80319; IGTM.
 DR InterPro: IPR001625; GLFV_dehydrog.
 DR Pfam: PF00208; GLFV_dehydrog; 1.

DR Pfam: PF02812; GLFV_dehydrog_N; 1.
 DR PRINTS: PR00082; GLFVDHGNASE.
 DR PROSITE: PS00074; GLFV_DEHYDROGENASE; 1.
 KM Oxidoreductase; NAD: Branched-chain amino acid metabolism.
 FT ACT_SITE 82 82 BY SIMILARITY.
 FT NP_BIND 182 188 NAD (POTENTIAL).
 SQ SEQUENCE 366 AA; 39867 MW; DA84E58052E772AC CRC64;

Query Match 63.5%; Score 33; DB 1; Length 366;
 Best Local Similarity 75.0%; Pred. No. 35;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSTVDSMD 9
 |::|||:
 Db 119 GTTVDDMD 126

RESULT 10
 ECB2_HALEL STANDARD; PRT; 421 AA.
 ID ECB2_HALEL
 AC O52250;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Diaminobutyrate-pyruvate aminotransferase (EC 2.6.1.46) (L-
 DE diaminobutyric acid transaminase) (Diaminobutyrate transaminase).
 GN ECTP.
 OS Halomonas elongata.
 OC Bacteria; Proteobacteria; gamma subdivision; Halomonadaceae;
 OC Halomonas.
 OX NCBI_TaxID=2746;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 25817;
 RX MEDLINE=98231640; PubMed=9570121;
 RA Goller K., Ofer A., Galinski E.A.;
 RT "Construction and characterization of an NaCl-sensitive mutant of
 RT Halomonas elongata impaired in ectoine biosynthesis."
 RL FEMS Microbiol. Lett. 161:293-300(1998).
 CC -I- CATALYTIC ACTIVITY: L-2,4-diaminobutanoate + pyruvate = L-
 CC aspartate 4-semialdehyde + L-alanine.
 CC -I- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -I- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF ECTOINE (1,4,5,6-
 CC TETRAHYDRO-2-METHYL-4-PYRIMIDINE CARBOXYLIC ACID).
 CC -I- SUBUNIT: HOMOHXAMER (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF031489; AAC15882.1; -
 DR InterPro: IPR00954; Aminoctran_3.
 DR Pfam: PF00202; amlnoctran_3; 2.
 DR PROSITE: PS00600; AA_TRANSFERRER_CLASS_3; FALSE NEG.
 KM Transferase; Aminotransferase; Pyridoxal phosphate.
 FT BINDING 267 267 PYRIDOXAL PHOSPHATE (POTENTIAL).
 SQ SEQUENCE 421 AA; 46166 MW; AAA2E21596E1E16C CRC64;

Query Match 63.5%; Score 33; DB 1; Length 421;
 Best Local Similarity 55.6%; Pred. No. 41;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSTVDSMDY 10
 ||::|:|:
 Db 176 GSSTDTLDY 184

```

RESULT 11
YRVA_CAEEL STANDARD: PRT: 518 AA.
AC 027519;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative cytochrome P450 CYP13A7 (EC 1.14.-.-).
GN CYP13A7 OR T10B9.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Gardner A.;
RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z48717; CAA8609.1; -
DR WormPep: T10B9.10; CE01655.
DR InterPro: IPR001128; Cyt_P450.
DR Pfam: PF00067; P450.1.
DR PRINTS: PR00385; P450.1.
DR PROSITE: PS00086; CYTOCHROME_P450.1.
DR KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
FT BINDING 464 464 HEME (BY SIMILARITY).
SQ SEQUENCE 518 AA; 58999 MW; CC04283EF87B9DEA7 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 518;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
DB 165 EGKTLDMLEY 174

RESULT 12
E13B_BACCI STANDARD: PRT: 682 AA.
AC P23903;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucan endo-1,3-beta-glucosidase A1 precursor (EC 3.2.1.39) ((1->3)-
DE beta-glucan endohydrolase) ((1->3)-beta-glucanase A1).
GN GlcA.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 39-52.
RC STRAIN-WL-12;
RC MEDLINE=90185240; PubMed=2311931;
RA Yahata N.; Watanabe T.; Nakamura Y.; Yamamoto Y.; Kamimiyu S.;
RA Tanaka H.;

```

```

RT "Structure of the gene encoding beta-1,3-glucanase A1 of Bacillus
RT circulans WL-12."
RL Gene 86:113-117(1990).
CC -1- FUNCTION: LYSIS OF CELLULAR WALLS CONTAINING BETA-1,3-GLUCANS.
CC IMPLICATED IN THE DEFENSE AGAINST FUNGAL PATHOGENS.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,3-beta-D-glucosidic linkages
CC in 1,3-beta-D-glucans.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M34503; AAA22474.1; -
DR PIR: J00420; J00420.
DR HSSP: P23904; IAOJ.
DR InterPro: IPR000757; Glyco_hydro.16.
DR Pfam: PF00722; Glyco_hydro.16; 1.
DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Cell wall; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 38
FT CHAIN 1 38
FT ACT_SITE 39 682 GLUCAN ENDO-1,3-BETA-GLUCOSIDASE A1.
FT ACT_SITE 552 552 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 557 557 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 682 AA; 75465 MW; 8CAF407E34D4ADD5 CRC64;

```

```

Query Match 63.5%; Score 33; DB 1; Length 682;
Best Local Similarity 66.7%; Pred. No. 71;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10
DB 40 GTTVISMEX 48

RESULT 13
PDC2_CANAL STANDARD: PRT: 836 AA.
ID PDC2_CANAL
AC 060035;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PDC2 protein.
GN PDC2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA Kaiser B.; Munder T.; Kuenkel W.; Eck R.;
RT "Molecular cloning and characterization of PDC2 from Candida
RT albicans."
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL FOR THE SYNTHESIS OF PYRUVATE DECARBOXYLASE
CC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y17007; CAA76580.1; -
FT DOMAIN 684 698 POLY-GLN.
FT DOMAIN 747 756 POLY-PRO.

```

SEQUENCE 836 AA; 94360 MW; EA030FC8D5C694C7 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 836;
Best Local Similarity 62.5%; Pred. No. 89;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

3 STVDSMDY 10
11:111
Db 817 STIDLDY 824

RESULT 14
ID CA16_MOUSE STANDARD: PRT: 1025 AA.
AC 004857;

DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(VI) chain precursor.
GN COL6A1.

OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

SEQUENCE FROM N.A.

RA MEDLINE=93316904; PubMed=8326912;
RA Bonaldo P., Piccolo S., Marvulli D., Volpin D., Bressan G.M.;
RT "Murine alpha 1(VI) collagen chain. Complete amino acid sequence and
RT identification of the gene promoter region.";
RL Matrix 13:223-233(1993).

RP MEDLINE=93256888; PubMed=8469506;
RX Zhang R.Z., Pan T.C., Timpl R., Chu M.L.;

RT "Cloning and sequence analysis of cDNAs encoding the alpha 1, alpha 2
RT and alpha 3 chains of mouse collagen VI.";
RL Biochem. J. 291:787-792(1993).

CC -1- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.

CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(VI),
CC ALPHA 2(VI), AND ALPHA 3(VI).

CC -1- P1M: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC -1- SIMILARITY: CONTAINS 3 VWFA DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: X66405; CAA47032.1; -;
DR EMBL: X66406; CAA47033.1; -;
DR EMBL: Z18271; CAA79152.1; -;

DR PIR: S34839; S34839.
DR PIR: S31403; S31403.
DR MGI: 88459; Col6a1.

DR InterPro: IPR000087; Collagen.
DR InterPro: IPR002035; VWFA.
DR Pfam: PF01391; Collagen; 5.

DR Pfam: PF00092; vwa; 3.
DR PRINTS: PR00453; VWFADOMAIN.
DR SMART: SM00327; VWA; 3.

DR PROSITE: PS50234; VWFA; 3.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Cell adhesion; Signal.

FT SIGNAL 1 19
FT CHAIN 20 1025 COLLAGEN ALPHA 1(VI) CHAIN.
FT DOMAIN 20 255 N-TERMINAL GLOBULAR DOMAIN.
FT DOMAIN 256 591 TRIPLE-HELICAL REGION.

FT DOMAIN 592 1025 C-TERMINAL GLOBULAR DOMAIN.
FT DOMAIN 36 234 VWFA 1.
FT DOMAIN 614 802 VWFA 1.
FT DOMAIN 826 1018 VWFA 3.
FT SITE 261 263 CELL ATTACHMENT SITE.
FT SITE 441 443 CELL ATTACHMENT SITE.
FT SITE 477 479 CELL ATTACHMENT SITE.
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 893 893 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 674 675 DM -> TL (IN REF. 2).
FT CONFLICT 709 709 T -> A (IN REF. 2).
FT CONFLICT 943 943 MISSING (IN REF. 2).
FT CONFLICT 960 960 O -> R (IN REF. 2).
SQ SEQUENCE 1025 AA; 108489 MW; 2A05DFED8771BBF7 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 1025;
Best Local Similarity 75.0%; Pred. No. 1,le+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

3 STVDSMDY 10
1:11111
Db 899 SSVDSMDF 906

RESULT 15
ID ACYO_MOUSE STANDARD: PRT: 98 AA.
AC P56376;

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acylphosphatase, organ-common type isozyme (EC 3.6.1.7)
DE (Acylphosphate phosphohydrolase).
GN ACYPL OR ACYPE.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

SEQUENCE FROM N.A.

RA Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,
RA Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
RA Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,
RA Theising B., Wylie T., Lennon G., Soares B., Wilson R., Waterston R.;

RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ITS PHYSIOLOGICAL ROLE IS NOT YET CLEAR.
CC -1- CATALYTIC ACTIVITY: An acyl phosphate + H(2)O -> a fatty acid anion
CC + phosphate.

CC -1- SIMILARITY: BELONGS TO THE ACYLPHOSPHATASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: AA142466; -; NOT ANNOTATED CDS.
DR EMBL: W83797; -; NOT_ANNOTATED_CDS.
DR HSP: P41500; 2ACY.

DR InterPro: IPR001792; Acylphosphatase.
DR Pfam: PF00708; Acylphosphatase; 1.
DR PRINTS: PR00112; ACYLPHPTASE.

DR Prodom: PD001884; Acylphosphatase; 1.
DR PROSITE: PS00150; ACYLPHOSPHATASE_1; 1.
DR PROSITE: PS00151; ACYLPHOSPHATASE_2; 1.
KW Hydroxylase; Acetylation; Multigene family.

FT INIT_MET 0 0 BY SIMILARITY.

FT	MOD_RES	1	1	ACETYLATION (BY SIMILARITY).
FT	ACT_SITE	23	23	POTENTIAL.
FT	ACT_SITE	41	41	POTENTIAL.
SO	SEQUENCE	98 AA;	11110 MW;	A73893A2B6AA8911 CRC64;

Query Match 61.5%; Score 32; DB 1; Length 98;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy	1	EGSTVDSMDY	10
	11	11	11
Db	2	EGDTLVSVDP	11

Search completed: August 20, 2002, 13:17:52
 Job time: 261 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:25 ; Search time 73.7 Seconds

(without alignments)
23.473 Million cell updates/sec

Title: US-09-824-286-3_COPY_95_104
Perfect score: 52
Sequence: 1 EGSTVDSMDY 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	75.0	5157	3	001135	001135 metarhizium
2	38	73.1	262	16	097019	097019 streptococc
3	38	73.1	345	5	09VXN9	09VXN9 drosophila
4	38	73.1	372	10	094H16	094H16 oryza sativ
5	37	71.2	329	16	098108	098108 rhizobium
6	37	71.2	1956	4	09Y5Y9	09Y5Y9 homo sapien
7	36	69.2	2005	5	09VPH5	09VPH5 drosophila
8	35	67.3	260	10	09STP2	09STP2 arabidopsi
9	35	67.3	280	5	018653	018653 caenorhabdi
10	35	67.3	484	5	09TZK6	09TZK6 caenorhabdi
11	35	67.3	535	2	068212	068212 actinomyces
12	35	67.3	544	4	09H087	09H087 homo sapien
13	35	67.3	621	4	094952	094952 homo sapien
14	35	67.3	1254	5	09VFS1	09VFS1 drosophila
15	34	65.4	117	2	09L4C2	09L4C2 uncultured
16	34	65.4	117	2	09L4C1	09L4C1 uncultured

17	34	65.4	117	2	09L4C0	09L4C0 uncultured
18	34	65.4	117	2	09L4B8	09L4B8 uncultured
19	34	65.4	117	2	09L4B6	09L4B6 uncultured
20	34	65.4	117	2	09L4B5	09L4B5 uncultured
21	34	65.4	117	2	09L4B4	09L4B4 uncultured
22	34	65.4	117	2	09L4B3	09L4B3 uncultured
23	34	65.4	117	2	09L4B1	09L4B1 uncultured
24	34	65.4	117	2	09L4A9	09L4A9 uncultured
25	34	65.4	117	2	09L4A8	09L4A8 uncultured
26	34	65.4	117	2	09L4A7	09L4A7 uncultured
27	34	65.4	117	2	09L4A6	09L4A6 uncultured
28	34	65.4	117	2	09L4A2	09L4A2 uncultured
29	34	65.4	117	2	09L4A1	09L4A1 uncultured
30	34	65.4	117	2	09L4A0	09L4A0 uncultured
31	34	65.4	117	2	09L4A9	09L4A9 uncultured
32	34	65.4	117	2	09L4A8	09L4A8 uncultured
33	34	65.4	233	2	093T99	093T99 azospirillum
34	34	65.4	233	2	093T97	093T97 azospirillum
35	34	65.4	257	16	09K0A6	09K0A6 vibrio chol
36	34	65.4	274	17	09YDH9	09YDH9 aeropyrum p
37	34	65.4	281	2	09ADV7	09ADV7 ehrlichia c
38	34	65.4	329	5	09TY08	09TY08 caenorhabdi
39	34	65.4	387	4	09HE7	09HE7 homo sapien
40	34	65.4	406	4	09H678	09H678 homo sapien
41	34	65.4	521	4	096SP3	096SP3 homo sapien
42	34	65.4	544	10	09FIF7	09FIF7 arabidopsis
43	34	65.4	831	4	09NVY8	09NVY8 homo sapien
44	34	65.4	896	4	09BVJ1	09BVJ1 homo sapien
45	33.5	64.4	303	16	097RB6	097RB6 streptococc

ALIGNMENTS

RESULT 1
ID 001135 PRELIMINARY: PRT; 5157 AA.
AC 001135;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PEPTIDE SYNTHETASE.
GN PESA.
OS Metarhizium anisopliae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocerales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
OX NCBI_TaxID=5350;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MEL;
RX MEDLINE=97082966; PubMed=8964498;
RA Bailey A.M., Kershaw M.J., Hunt B.A., Patterson I.C., Charnley A.K.,
Raynolds S.E., Clarkson J.M.;
RT "Cloning and sequence analysis of an intron containing domain of a
peptide synthetase from the entomopathogenic fungus Metarhizium
anisopliae.";
RT Gene 173:195-197(1996).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=MEL;
RA Bailey A.M., Reynolds S.E., Charnley A.K., Clarkson J.M.;
RT "Evidence for multiple peptide synthetases from Metarhizium
anisopliae.";
RT Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; X89442; CAA61605.1; -;
DR HSBP; P14687; IAMU.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; DUF4.
DR InterPro: IPR003880; Phosphonat_attach.
DR Pfam: PF00501; AMP-binding; 4.
DR Pfam: PF00668; Condensation; 7.
DR Pfam: PF00550; PP-binding; 4.
DR PRINTS; PR00154; AMPBINDING.

DR PROSITE: PS50075; ACP_DOMAIN; 4.
 DR PROSITE: PS00455; AMP_BINDING; 4.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
 KW Phosphopantetheine.
 SQ SEQUENCE 5157 AA; 573954 MW; 1038242BA3143668 CRC64;

Query Match 75.0%; Score 39; DB 3; Length 5157;
 Best Local Similarity 77.8%; Pred. No. 2.4e+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGSTVDSMD 9
 DB 2003 DGVSDSLD 2011

RESULT 2
 ID 097019 PRELIMINARY; PRT: 262 AA.
 AC 097019;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PROIPOPOTEIN DIACYLGLYCERYL TRANSFERASE.
 GN SP1412.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayahara E., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouli H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougethly B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae."
 RT Science 293:498-506(2001).
 RL EMBL: AE007438; AAK75510.1; -.
 DR TIGR: SP1412; -.
 DR InterPro: IPR001640; LGT.
 DR Pfam: PF01790; LGT; 1.
 DR PROSITE: PS01311; LGT; 1.
 KW Transferase; Lipoprotein; Complete proteome.
 SQ SEQUENCE 262 AA; 30292 MW; B3A58C5F4BDC19 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 262;
 Best Local Similarity 66.7%; Pred. No. 15;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSTVDSMDY 10
 DB 142 GATVDNLDY 150

RESULT 3
 ID 09VXN9 PRELIMINARY; PRT: 345 AA.
 AC 09VXN9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CG895 PROTEIN (PEPTIDOGLYCAN-RECOGNITION PROTEIN-LE) (GH01554P).
 GN PGRP-LE OR CG895.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhendari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burlis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Zhang G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RT Science 287:2185-2195(2000).
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20558582; PubMed=11106397;
 RA Werner T., Liu G., Kang D., Ekengren S., Steiner H., Hultmark D.,
 RT "A family of peptidoglycan recognition proteins in the fruit fly
 Drosophila melanogaster."
 RT Proc. Natl. Acad. Sci. U.S.A. 97:13772-13777(2000).
 RL [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Y, CN BW SP;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Chame M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno C., Paclab J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBI databases.
 DR EMBL: AE003500; AAF48519.1; -.
 DR EMBL: AF313391; AAG32064.1; -.
 DR EMBL: AY058258; AAL15487.1; -.
 DR Flybase: FBgn0030695; PGRP-LE.
 SQ SEQUENCE 345 AA; 39426 MW; 66CC484B54705AD7 CRC64;

Query Match 73.1%; Score 38; DB 5; Length 345;
 Best Local Similarity 70.0%; Pred. No. 20;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10

Db 1 1 111:11
31 ESSYVDSLDY 40

RESULT 4
ID Q94HI6 PRELIMINARY; PRT: 372 AA.
AC Q94HI6:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEtical PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang R.A., Fisch D., Presting G., Wood T., Yu Y., Soderlund C.,
RA Kim H.-R., Rambo T., Henry D., Simmons J.;
RT "Rice Genomic Sequence."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC079632; AAK63901.1; -
SQ SEQUENCE 372 AA; 42154 MW; D9F3C33822D5A2AB CRC64;

Query Match 73.1%; Score 38; DB 10; Length 372;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGSTVDSMD 9
Db 29 EGATIDNMD 37

RESULT 5
ID Q981U8 PRELIMINARY; PRT: 329 AA.
AC Q981U8:
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE MLP228 PROTEIN.
GN MLP228.
OS Rhizobium loti (Mesorhizobium loti).
OC Grammid pmla.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Katano A., Idesawa K., Ishikawa K., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003015; BAB54611.1; -
KW Plasmid; Complete proteome.
SQ SEQUENCE 329 AA; 35675 MW; 86CEACBBI55345DC CRC64;

Query Match 71.2%; Score 37; DB 16; Length 329;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
1 11111111

Db 207 EGSTVETVDY 216

RESULT 6
ID Q9Y5Y9 PRELIMINARY; PRT: 1956 AA.
AC Q9Y5Y9:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TETRODOTOXIN-RESISTANT VOLTAGE-GATED SODIUM CHANNEL.
GN SCN10A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=DORSAL ROOT GANGLION;
RX MEDLINE=9905355; PubMed=9839820;
RA Robert D.K., Koch B.D., Illicka M., Obernolte R.A., Naylor S.L.,
RA Herman R.C., Eglen R.M., Hunter J.C., Sangameswaran L.;
RT "A tetrodotoxin-resistant voltage-gated sodium channel from human
dorsal root ganglia, hPNS/SCN10A."
RL Pain 78:107-114(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=DORSAL ROOT GANGLION;
RA Robert D.K., Koch B.D., Illicka M., Obernolte R.A., Naylor S.L.,
RA Herman R.C., Eglen R.M., Hunter J.C., Sangameswaran L.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF117907; AAD30863.1; -
DR InterPro: IPR000636; Catlon_chan_non_lig.
DR InterPro: IPR002111; Cat_channel_TrpL.
DR InterPro: IPR001682; Channel_pore_Ca_Na.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR001696; Na_channel.
DR Pfam: PF00520; Ion_Trans; 4.
DR Pfam: PF00612; IQ; 1.
DR PRINTS: PR00170; MACHANNEL.
KW Ionic channel.
SQ SEQUENCE 1956 AA; 220564 MW; F24B73E7C211DA5E CRC64;

Query Match 71.2%; Score 37; DB 4; Length 1956;
Best Local Similarity 77.8%; Pred. No. 21e02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGSTVDSMD 9
Db 1085 EGSTVDCID 1093

RESULT 7
ID Q9VFH5 PRELIMINARY; PRT: 2005 AA.
AC Q9VFH5:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE CADHERIN PRECURSOR (CG3389 PROTEIN).
GN CG3389.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Suttton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazei R.G., Chame M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hostin D., Houston K.A., Howland T.J., Mel M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
 Lasko F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Splier E., Spradling A.C., Stapleton R., Strong R., Sun E.,
 RA Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*.";
 Science 287:2185-2195(2000).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
 CC EMBL: AE003705; AAF5082.1; -.
 DR HSPSP; P15116; INCI.
 DR FlyBase: FBgn0038247; CG3389.
 DR InterPro: IPR002126; Cadherin.
 DR Pfam: PF00028; cadherin; 14.
 DR PRINTS: PR00205; CADHERIN.
 DR SMART: SM00112; CA; 14.
 DR PROSITE: PS00232; CADHERIN_1; 1.
 DR PROSITE: PS00268; CADHERIN_2; 14.
 DR KW Hypothetical protein; Cell adhesion; signal; Transmembrane; Repeat;
 KW Glycoprotein; Calcium-binding
 FT SIGNAL 1 55 POTENTIAL.
 FT CHAIN 56 2005 POTATIVE CADHERIN.
 FT DOMAIN 56 1849 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1850 1870 POTENTIAL.
 FT DOMAIN 1871 2005 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 75 172 CADHERIN 1.
 FT DOMAIN 181 281 CADHERIN 2.
 FT DOMAIN 290 393 CADHERIN 3.
 FT DOMAIN 402 510 CADHERIN 4.
 FT DOMAIN 519 621 CADHERIN 5.
 FT DOMAIN 630 746 CADHERIN 6.
 FT DOMAIN 755 880 CADHERIN 7.
 FT DOMAIN 888 1001 CADHERIN 8.
 FT DOMAIN 1010 1105 CADHERIN 9.
 FT DOMAIN 1114 1219 CADHERIN 10.
 FT DOMAIN 1228 1328 CADHERIN 11.
 FT DOMAIN 1339 1438 CADHERIN 12.
 FT DOMAIN 1447 1629 CADHERIN 13.
 FT DOMAIN 1638 1743 CADHERIN 14.
 FT CARBOHYD 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 516 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 516 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 867 867 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1031 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1072 1072 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1111 1111 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1293 1293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1334 1334 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1342 1342 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1446 1446 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1516 1516 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1529 1529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1718 1718 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1764 1764 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 2005 AA; 224474 MW; C3616F442AFADDE2 CRC64;

Query Match Score 69.2%; DB 5; Length 2005;
 Best Local Similarity 66.7%; Pred No. 3.5e-02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10
 DB 11 GNGIDSMXD 19

RESULT 8
 ID Q9STP2 PRELIMINARY; PRT; 260 AA.
 AC Q9STP2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 13, Last annotation update)
 DE HYPOTHETICAL 29.2 KDA PROTEIN.
 GN T27E11.80 OR AT4G27840.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Arabidopsis: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicot: Rosidae;
 OC Eurosids II: Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_Taxid=3702;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,
 RA Barrell B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K.,
 RA Schellier C.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [13]
 RP SEQUENCE FROM N.A.
 RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [14]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL078579; CAB43972.1; -.
 DR EMBL: AL161571; CAB41433.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 260 AA; 29203 MW; 6482A1D19C3A3C47 CRC64;

Query Match Score 67.3%; DB 10; Length 260;
 Best Local Similarity 66.7%; Pred No. 59;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMD 9
 DB 98 DSTADSLD 106


```

RESULT 9
ID 018663 PRELIMINARY; PRT: 280 AA.
AC 018663;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOTHETICAL 31.5 KDA PROTEIN.
GN C46H3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Favello A.;
RT "The sequence of C. elegans cosmid C46H3.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL EMBL: U01271; AAA82445.1; -.
DR InterPro:IPR002110; ANK.
DR Pfam: PF00023; ank; 6.
DR SMART: SM00248; ANK; 5.
DR PROSITE: PS50088; ANK_REPEAT; 3.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
KW ANK repeat; Hypothetical protein; Repeat.
SQ SEQUENCE 280 AA; 31530 MW; 254EBA4EDE033866 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 280;
Best Local Similarity 66.7%; Pred. No. 64;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSTVDSMDY 10
Db 130 GANVDSVDY 138
1:|||||
|:|||||

RESULT 10
ID 097ZK6 PRELIMINARY; PRT: 484 AA.
AC 097ZK6;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE C36C9.5 PROTEIN.
GN C36C9.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

```

```

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton B., Martin J., O'Brien D.;
RT "The sequence of C. elegans cosmid C36C9.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF098986; AAC67427.1; -.
SQ SEQUENCE 484 AA; 55327 MW; CCC1ECBE82CA441 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 484;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
Db 235 EGSSSETMDY 244
|||:||||
|:|||||

RESULT 11
ID 068212 PRELIMINARY; PRT: 535 AA.
AC 068212;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE FIMBRIAL STRUCTURAL SUBUNIT.
GN FIMA.
OS Actinomyces naeslundii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetes.
OX NCBI_Taxid=1655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T14V;
RX MEDLINE=98187920; PubMed=9529071;
RA Yeung M.K., Donkersloot J.A., Cisar J.O., Ragsdale P.A.;
RT "Identification of a gene involved in assembly of Actinomyces
RT naeslundii T14V type 2 fimbriae.";
RL Infect. Immun. 66:1482-1491(1998).
DR EMBL: AF019629; AAC13545.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN; 1.
SQ SEQUENCE 535 AA; 56038 MW; B115867F500C1356 CRC64;

Query Match 67.3%; Score 35; DB 2; Length 535;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
Db 263 EGSTLDPTDY 272
|||:||||
|:|||||

RESULT 12
ID 09H087 PRELIMINARY; PRT: 544 AA.

```

```
AC Q9H087;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEICAL 63.3 KDA PROTEIN.
GN DKFZP346G058.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA MEDLINE=21154917; PubMed=11230166;
RA Wilmann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.W., Ottenwaelder B., Obermayer B., Tampe J., Heubner D.,
RA Wandt R., Korn B., Klein M., Poustka A.;
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL EMBL; AL136899; CAB66833.1; -.
KW Hypothetical protein.
SQ SEQUENCE 544 AA; 63306 MW; D95BE3CEA3B7D0ED CRC64;

Query Match 67.3%; Score 35; DB 4; Length 544;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
DB 253 EGATLIDIFY 262

RESULT 13
Q94952 PRELIMINARY; PRT; 621 AA.
ID 094952;
AC 094952;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE KIA0875 PROTEIN (FRAGMENT).
GN KIA0875 OR FBX21.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RA MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=20003060; PubMed=10531035;
RA Cenciarelli C., Chatur D.S., Guardavaccaro D., Parks W., Vidal M.,
RA Pagano M.;
RT "Identification of a family of human F-box proteins.";
RL Curr. Biol. 9:1177-1179(1999).
DR EMBL; AB020682; BAA74898.1; -.
DR EMBL; AF174601; AAF04522.1; -.
DR InterPro; IPR001810; F-box.
DR Pfam; PF00646; F-box; 1.
FT NON_TER 1
SQ SEQUENCE 621 AA; 71479 MW; D302DD46F3ABB5A3 CRC64;
```

```
Query Match 67.3%; Score 35; DB 4; Length 621;
Best Local Similarity 60.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
DB 337 EGATLIDIFY 346

RESULT 14
Q9VF51 PRELIMINARY; PRT; 1254 AA.
ID 09VF51;
AC 09VF51;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE CG6045 PROTEIN.
GN CG6045.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burlits K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J.D., Puri V., Reese M.G.,
RA Rehner K., Remington K., Saunders R.D.C., Schelder F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtikas R., Tector C., Turner R., Venter E., Wang A., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodde T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -i- COFACTOR: BINDS A 2PE-2S CLUSTER (BY SIMILARITY).
DR EMBL; AB003709; AAF55209.1; -.
DR FlyBase; FBgn0038349; CG6045.
DR InterPro; IPR002888; 2Pe-2S_BD.
DR InterPro; IPR000564; 2Pe2S_ferredoxin.
DR InterPro; IPR000674; Aldxan_dh.C.
DR InterPro; IPR002346; dehydrog_mol_yb.
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:24:06 ; Search time 49.92 seconds

(without alignments)
24.475 Million cell updates/sec

Title: US-09-824-286-4_COPY_24_34

Sequence: 1 KASQDVTAVA 11

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

A_Geneseq_032802:*

1: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	50	100.0	106	19	AAW31647	Monoclonal antibod
2	46	92.0	250	17	AAW02278	741F8 anti-c-erbB-
3	46	92.0	250	18	AAW29261	Anti-c-erbB-2 sing
4	46	92.0	250	18	AAW22400	Single chain antib
5	46	92.0	250	19	AAW53168	741F8 anti-c-erbB-
6	46	92.0	250	19	AAW47012	Single chain bindi
7	46	92.0	250	20	AAW80432	Anti-c-erbB-2 sFv
8	46	92.0	667	14	AAW39573	Sequence of 741 sF
9	45	90.0	11	16	AAW4946	L-CDR-1 of anti-id
10	45	90.0	11	19	AAW62189	Mouse anti-HM1.24
11	45	90.0	11	19	AAW57577	Chimeric L chain V

12	45	90.0	11	20	AAW02547	Artificial CDR(1)
13	45	90.0	11	20	AAW89629	Mouse humanised an
14	45	90.0	11	21	AAW77515	Antibody H chain V
15	45	90.0	11	22	AAW67113	Amino acid sequenc
16	45	90.0	11	22	AAW63395	Amino acid sequenc
17	45	90.0	11	22	AAW64777	Anti-PTHRP Ab VL C
18	45	90.0	11	22	AAW76891	Human PTHrP mouse
19	45	90.0	11	22	AAW76910	Human PTHrP mouse
20	45	90.0	11	22	AAW76929	Human PTHrP mouse
21	45	90.0	107	16	AAW74958	Immunoglobulin lig
22	45	90.0	109	14	AAW30766	Light chain variab
23	45	90.0	126	19	AAW77293	HM1.24 antibody 11
24	45	90.0	126	19	AAW65772	Anti-human HM1.24
25	45	90.0	126	19	AAW62208	Humanised anti-HM1
26	45	90.0	126	19	AAW62195	Human anti-HM1.24
27	45	90.0	126	20	AAW32769	Anti-HM1.24 antibo
28	45	90.0	126	20	AAW02553	Humanised L chain
29	45	90.0	126	20	AAW02554	Humanised L chain
30	45	90.0	126	20	AAW05485	H chain V region o
31	45	90.0	130	16	AAW74968	Anti-idiotypic anti
32	45	90.0	131	19	AAW62187	Mouse anti-HM1.24
33	45	90.0	131	20	AAW02545	L chain V region o
34	45	90.0	11	16	AAW85774	L-CDR-1 of anti-id
35	44	88.0	107	16	AAW74959	Immunoglobulin lig
36	44	88.0	120	16	AAW74969	Anti-idiotypic anti
37	43	86.0	108	20	AAW21817	Anti-STX1 light ch
38	42	84.0	11	19	AAW47101	NMDA receptor bind
39	42	84.0	109	14	AAW30762	Light chain variab
40	42	84.0	109	15	AAW47039	Sequence of the 11
41	42	84.0	237	13	AAW24047	Light chain of 4D5
42	42	84.0	252	21	AAW28158	Human anti-c-ErbB2
43	42	84.0	513	22	AAW82924	Anti-HER2 Humab4D5
44	42	84.0	698	20	AAW83493	4D5 Fab molecule e
45	40	80.0	11	17	AAW97313	Humanised monoclona

ALIGNMENTS

RESULT 1	
ID	AAW31647 standard; Protein; 106 AA.
XX	
AC	AAW31647;
XX	
DT	21-MAY-1998 (first entry)
XX	
DE	Monoclonal antibody CP.B8 light chain variable region.
XX	
KW	Cytokine receptor; gamma common chain; gc chain; human;
KW	blocking agent; monoclonal antibody; CP.B8; immunological disease;;
KW	myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
KW	insulin-dependent diabetes; inflammatory bowel disease;
KW	sympathetic ophthalmia; uveitis; allergy; asthma; infection;;
KW	grit versus host disease; psoriasis; immunosuppressive; therapy;
KW	complementarily determining region; CDR.
XX	
OS	Mus musculus.
XX	
PH	Key
FT	Region
FT	Location/Qualifiers
FT	24..34
FT	/note= "CDR1"
FT	50..56
FT	/note= "CDR2"
FT	89..97
FT	/note= "CDR3"
XX	
PN	MO9743416-AL.
XX	
PD	20-NOV-1997.
XX	
PF	09-MAY-1997; 97WO-US07870.
XX	

10-MAY-1996; 96US-0017466.
 (BIOI) BIOGEN INC.
 Benjamin CD, Burkly LC, Hession C, Whitty A;
 WPI; 1998-008885/01.
 N-PSDB; AAT97440.
 Blocking agents of the gamma common chain of cytokine receptors -
 particularly monoclonal antibodies, used to induce T cell anergy for
 treatment of immunological diseases
 Claim 22; Page 81; 11np; English.
 This polypeptide comprises the light chain variable region (VL) of
 monoclonal antibody (MAb) C9.88, which is produced by a hybridoma
 deposited as ATCC 12107, and which is specific for the gamma
 constant (gc) chain (see AAMJ1646) of human cytokine receptors. The
 invention provides compositions and methods for inhibiting cytokine
 signalling using gc chain blocking agents for the treatment of
 immunological diseases such as myasthenia gravis, rheumatoid
 arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,
 inflammatory bowel disease, sympathetic ophthalmia, uveitis,
 allergy, asthma, parasitic infection, graft vs. host disease, or
 psoriasis. Preferred gc blocking agents include MAb CP.88, its Fab
 fragment and an antibody having a light chain variable region
 CDR selected from those of CP.88 VL or a heavy chain variable
 region CDR selected from those of CP.88 VH (see AAMJ1648).

Query Match	100.0%	Score 50;	DB 19;	Length 106;
Best Local Similarity	100.0%	Pred. No. 0.057;		
Matches 11; Conservative	0;	Mismatches	0;	Indels
				Gaps 0;
QY	1	KASQDVTYTA	11	
DB	24	KASQDVTYTA	34	

RESULT	2	
AA002278		
AA002278	standard; Protein; 250 AA.	
XX		
AC	AA002278:	
XX		
29-OCT-1996	(first entry)	
XX		
DE	7A1F8 anti-c-erbB-2 two single chain Fv construct.	
XX		
7A1F8:	anti-c-erbB-2; monoclonal antibody; single chain Fv; sfv construct; polypeptide linker; C-terminal amino acid sequence; in vivo imaging; drug targeting experiment; homodimer; increased; binding avidity; tissue retention time.	
KM		
XX		
CS	Homo sapiens.	
XX		
Key	Location/Qualifiers	
FH	122..135	
FT	/label= linker	
FT	246..250	
Peptide	/note="Claimed C-terminal tail to facilitate crosslinking of two sfv polypeptides"	
FT		
XX		
PN	US5534254-A.	
XX		
PD	09-JUL-1996.	
XX		
PF	06-FEB-1992; 92US-0831967.	
XX		
07-OCT-1993.:	93US-0133804.	
PR		

XX	06-FEB-1992:	92US-0831967.
XX	(CHIR) CHIRON CORP.	
PA	(CREA-) CREATIVE BIOMOLECULES INC.	
XX		
PI	Houston LL, Huston JS, Oppermann H, Ring DB;	
DR	WPI; 1996-333194/33.	
DR	N-PDDB; AAT36878.	
PT	Compsns. contg. antigen-targetting antibody fragment constructs	
XX	comprising dimer of single-chain Fv fragments	
PS	Claim 25; Columns 27-28; 30pp; English.	
XX		
CC	The variable heavy (VH) and variable light (VL) genes of the 7A1F8	
CC	anti-C-eIbB-2 monoclonal antibody (MAb), were isolated from the	
CC	cDNA of the parental 7A1F8 hybridoma line. A two single chain Fv	
CC	(sfv) gene was constructed by connecting the VH and VL genes with a	
CC	DNA sequence encoding a polypeptide linker. A synthetic DNA duplex	
CC	encoding the C-terminal amino acid sequence, (gly) ⁴ -Cys was	
CC	inserted, and the resulting 7A1F8 anti-C-eIbB-2 two sfv inserted	
CC	into an expression vector. The resulting gene, which encodes the	
CC	present sequence, was transformed into E. coli, and protein	
CC	expression induced by the addn. of IPTG to the culture medium.	
CC	A compsn. comprising a carrier and the 2 sfv protein prod. can be	
CC	used for in vivo imaging, and drug targetting experiments. The	
CC	2 sfv protein prod. is a homodimer, in which both fragments target	
CC	the same antigen, therefore giving greater binding avidity and	
CC	longer tissue retention times, compared to individual sfv protein	
CC	prod. fragments.	
XX		
XQ	Sequence 250 AA;	

Query Match	92.0%;	Score 46;	DB 17;	Length 250;
Best Local Similarity	90.9%;	Pred. No. 0.75;		
Matches 10; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	KASODVTTAVA	11
Db	159	kasqdvstava	169
	RESULT	3	
	AAW29261		
ID	AAW29261	standard; Protein; 250 AA.	
AC	AAW29261;		
XX			
DT	07-JAN-1998	(first entry)	
XX			
DE	Anti-c-erbB-2 single chain antibody 741F8.		
KM	Adenovirus; E1A; transactivator; transcription activator; stimulate		
KM	expression vector; single-chain binding protein; VAI; enhance; PCR;		
KM	translation; production; immortal; eukaryotic cell; scfv; primer;		
KM	single-chain antibody fragment; imaging; tumour; breast cancer;		
KM	ovarian cancer; c-erbB-2 antigen; digoxin intoxication.		
XX			
OS	Synthetic.		
PN	US5658763-A.		
XX			
PD	19-AUG-1997.		
XX			
PE	25-OCT-1993;	93US-0143498.	
XX			
PR	25-OCT-1993;	93US-0143498.	
XX			
PA	05-JUN-1995;	95US-0463675.	
XX			
XX	(CREA-)	CREATIVE BIOMOLECULES INC.	

PI Doral H, Oppermann H;
 XX WPI: 1997-424235/39.
 DR N-PSDB; AAT91837.
 XX
 PT Producing single chain binding protein in immortalised eukaryotic
 PT cells - which comprise protein coding sequences, a transcription
 PT activator and translation promotion sequences, provides high
 PT expression at low copy number
 XX
 PS Example 2; Column 29-32; 24pp; English.
 XX
 CC This sequence is a single chain anti-c-erbB-2 antibody (scFv)
 CC derived from hybridoma 741F8. The protein product includes a C-terminal
 CC tail Gly4-Cys in some constructs. Single-chain antibodies can be
 CC produced using a novel method which comprises culturing an immortalised
 CC eukaryotic cell having transfected DNA sequences (encoding the protein
 CC of interest), integrated into its genome. In particular expression
 CC effector vectors containing a non-native reporter DNA (encoding the scFv)
 CC and viral sequences to promote transcription and translation (e.g. the
 CC adenovirus E1A and VAI genes as shown in AAT91831 and AAT91834
 CC respectively) are used. The scFv that is produced, when properly folded,
 CC has a structure with mono- or bi-functional binding activity. The method
 CC is especially used to produce single-chain antibody fragments (scFv),
 CC e.g. for imaging tumours or delivering therapeutic agents to them,
 CC particularly breast and ovarian cancers that express the c-erbB-2
 CC antigen. Other scFv are used in model studies and for treating digoxin
 CC intoxication.
 XX
 SQ Sequence 250 AA;
 XX
 Query Match 92.0%; Score 46; DB 18; Length 250;
 Best Local Similarity 90.9%; Pred. No. 0.75;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 OY 1 KASQDVTTAVA 11
 |||||:||||
 Db 159 kasqdvstava 169
 XX
 RESULT 4
 AAW22400
 ID AAW22400 standard; Protein; 250 AA.
 XX
 AC AAW22400;
 XX
 DT 09-OCT-1997 (first entry)
 XX
 DE Single chain antibody 741F8 protein sequence.
 XX
 KW Production: single-chain; binding protein; antibody; eukaryote; virus;
 KW transcription activator; promoter; expression; adenovirus; E1A; PCR;
 KW polymerase chain reaction; amplification; primer; herpes simplex virus;
 KW thymidine kinase; vector; enhancer; translation; heterologous.
 XX
 OS Synthetic.
 XX
 PN US5631158-A.
 XX
 PD 20-MAY-1997.
 XX
 PF 25-OCT-1993; 93US-0143498.
 XX
 PR 25-OCT-1993; 93US-0143498.
 PR 05-JUN-1995; 95US-0461184.
 XX
 PA (CREA-) CREATIVE BIOMOLECULES INC.
 PI Doral H, Oppermann H;
 XX
 WPI: 1997-288577/26.
 DR N-PSDB; AAT78879.

XX
 PT Production of cell line for producing single-chain binding protein -
 PT using construct containing DNA encoding viral transcription
 PT activator protein
 XX
 PS Example 2; Column 29-32; 24pp; English.
 XX
 CC The invention relates to methods of increasing production of a
 CC single-chain binding protein, especially a single chain antibody,
 CC by generating eukaryotic cell lines containing DNA encoding either
 CC a viral transcription activator protein that acts on and stimulates
 CC a viral promoter controlling the expression of DNA encoding the
 CC single-chain binding protein, such as the adenovirus Ad2 E1A protein
 CC encoded by the sequence AAT78873, or an RNA sequence able to promote
 CC translation of the RNA transcript from the heterologous gene, such as
 CC the adenoviral VAI gene (AAT78876).
 CC The sequence presented here is the amino acid sequence of the
 CC anti-c-erbB2 single chain antibody 741. The coding sequence was
 CC constructed by amplifying the variable heavy and light chain genes
 CC which were then cloned into a pUC vector. For secretion from mammalian
 CC cells, the sequence is preceded by the signal peptide sequence from
 CC the monoclonal antibody 520C9 (AAT78880) or from Pac1 (AAT78881).
 XX
 SQ Sequence 250 AA;
 XX
 Query Match 92.0%; Score 46; DB 18; Length 250;
 Best Local Similarity 90.9%; Pred. No. 0.75;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 OY 1 KASQDVTTAVA 11
 |||||:||||
 Db 159 kasqdvstava 169
 XX
 RESULT 5
 AAW53168
 ID AAW53168 standard; Protein; 250 AA.
 XX
 AC AAW53168;
 XX
 DT 16-JUL-1998 (first entry)
 XX
 DE 741F8 anti-c-erbB-2 sfv' dimeric construct protein sequence.
 XX
 KW Antigen imaging; single chain Fv; sfv; linker; dimeric; cancer;
 KW c-erbB-2; tumour; diagnosis.
 XX
 OS Synthetic.
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Peptide 122..135
 FT Peptide /note="Linker peptide"
 FT Peptide 246..250
 FT /note="Gly4-Cys C-terminal tail"
 XX
 EN US5753204-A.
 XX
 PD 19-MAY-1998.
 XX
 PF 05-JUN-1995; 95US-0461838.
 XX
 PR 07-OCT-1993; 93US-0133804.
 PR 06-FEB-1992; 92US-0831967.
 PR 05-JUN-1995; 95US-0461838.
 XX
 PA (CHIR) CHIRON CORP.
 PI (CREA-) CREATIVE BIOMOLECULES INC.
 XX
 Houston LL, Huston JS, Oppermann H, Ring DB;
 WPI: 1998-311318/27.

DR N-PSDB: AAV21796.
 XX
 XX Imaging of antigens in vivo - using dimers of single-chain antibody
 PT Fv fragments
 XX
 XX
 PS Example 1; Columns 25-28; 30pp; English.
 XX
 CC This represents the protein sequence of a 741F8 sfv' (single chain Fv)
 CC C-terminal Gly4-Cys construct. This was constructed by connecting the vh
 CC and vl genes with a DNA sequence encoding a 14 residue polypeptide
 CC linker. 741F8 is a monoclonal antibody useful in targeting C-erbB-2
 CC antigen. This dimeric construct can be used in the methods of invention
 CC of imaging a preslected antigen expressed in a mammal. The methods are
 CC used in magnetic resonance imaging of C-erbB-2 or related antigens in
 CC cancer diagnosis. The biosynthetic constructs have enhanced properties
 CC as in vivo targeting agents in comparison with intact monoclonal
 CC antibodies or their Fab fragments. The dimeric constructs permit the
 CC in vivo targeting of an epitope on an antigen with greater apparent
 CC avidity, including greater tumour specificity, tumour localisation and
 CC tumour retention properties than that of the Fab fragment having the same
 CC CDRs as the construct.
 CC
 XX Sequence 250 AA;
 SO
 Query Match 92.0%; Score 46; DB 19; Length 250;
 Best Local Similarity 90.9%; Pred. No. 0.75;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 OY 1 KASQDVTTAVA 11
 |||||:||||
 Db 159 kasqdvstava 169
 RESULT 6
 AAM47012
 ID AAM47012 standard; Protein: 250 AA.
 XX
 AC AAM47012;
 XX
 DT 13-JUL-1998 (first entry)
 XX
 DE Single chain binding site molecule (sfv) 741F8.
 XX
 KM Single chain binding site molecule 741F8; sfv 741F8; large scale;
 KM transactivating transcription activator; viral transcription promoter;
 KM protein production; non-native gene; hard to express gene.
 XX
 OS Synthetic.
 XX
 FN US5733782-A.
 XX
 PD 31-MAR-1998.
 XX
 PE 05-JUN-1995; 95US-0464589.
 XX
 PR 25-OCT-1993; 93US-0143498.
 XX
 PA (CREA-) CREATIVE BIOMOLECULES INC.
 XX
 PI Doral H, Oppermann H;
 XX
 DR WPI; 1998-229831/20.
 DR N-PSDB; AAV22398.
 XX
 PT Immobilised eukaryotic cell comprising transfected DNA sequences -
 PT useful for enhanced production of proteins encoded by non native
 PT genes
 XX
 PS Example 2; Columns 31-32; 24pp; English.
 XX
 CC The present sequence represents single chain binding site molecule (sfv).
 CC The sequence was expressed to exemplify the invention, which is a method

CC for the expression of poorly expressed genes in an immortalised
 CC eukaryotic cell. This cell contains transfected DNA sequences operatively
 CC integrated into its genome. The transfected DNA sequences encode a viral
 CC transcription promoter linked to a non-native reporter sequence encoding
 CC a single chain binding protein. The promoter is activated by a viral
 CC transcription activator protein. The cell allows the production on
 CC a commercial scale of proteins encoded by non-native hard to express
 CC genes.
 CC
 XX Sequence 250 AA;
 SO
 Query Match 92.0%; Score 46; DB 19; Length 250;
 Best Local Similarity 90.9%; Pred. No. 0.75;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 OY 1 KASQDVTTAVA 11
 |||||:||||
 Db 159 kasqdvstava 169
 RESULT 7
 AAM80422
 ID AAM80422 standard; Protein: 250 AA.
 XX
 AC AAM80422;
 XX
 DT 28-JAN-1999 (first entry)
 XX
 DE Anti-c-erbB-2 sfv' protein sequence.
 XX
 KM Anti-c-erbB-2 sfv'; c-erbB-2; antigen; tumour cell; antibody 741F8;
 KM targeted delivery; antigen-expressing cell.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 FT Key Location/Qualifiers
 FT Peptide 122..135
 FT /note="linker"
 XX
 PN US5837846-A.
 XX
 PD 17-NOV-1998.
 XX
 PE 05-JUN-1995; 95US-0461386.
 XX
 PR 07-OCT-1993; 93US-0133804.
 PR 06-FEB-1992; 92US-0831967.
 PR 05-JUN-1995; 95US-0461386.
 XX
 PA (CHIR) CHIRON CORP.
 PA (CREA-) CREATIVE BIOMOLECULES INC.
 XX
 PI Houston LL, Huston JS, Oppermann H, Ring DB;
 XX
 DR WPI; 1999-023541/02.
 DR N-PSDB; AAV63397.
 XX
 PT Nucleic acid encoding single-chain Fv fragment specific for antigens
 PT - and having C-terminal tail for crosslinking to form dimer with
 PT improved pharmacokinetic properties, used to deliver drugs and
 PT imaging agents, especially to tumours
 XX
 PS Example 1; Columns 27-30; 29pp; English.
 XX
 CC The present sequence represents an anti-c-erbB-2 sfv'. c-erbB-2 is an
 CC antigen that is overexpressed on the surface of tumour cells. An
 CC antibody designated 741F8 binds c-erbB-2. Variable heavy and light
 CC sequences of antibody 741F8 are connected, together with a linker, to
 CC produce the present single chain Fv gene. Anti-c-erbB-2 sfv' exemplifies
 CC the invention. Dimers of the single chain Fv are used for targeted
 CC delivery of drugs or imaging agents (e.g. cytotoxins, prodrugs or

CC 99m-technetium) to antigen-expressing cells, particularly for treatment
CC or diagnosis of tumours (especially of ovary or breast).

SO Sequence 250 AA;

Query Match 92.0%; Score 46; DB 20; Length 250;
Best Local Similarity 90.9%; Pred. No. 0.75;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTAVA 11
| | | | | | | | | |
DB 159 kasqdvstava 169

RESULT 8

AAAR39573
ID AAR39573 standard; Protein; 667 AA.

AC AAR39573;

DT 07-FEB-1994 (first entry)

DE Sequence of 741 SFV-PE40.

KW Single chain Fv polypeptide; VH-VL heterodimer; immunoglobulin; Ig;
exotoxin.

OS Pseudomonas.

PN MO9316185-A.

PD 19-AUG-1993;

PF 05-FEB-1993; 93WO-US01055.

PR 06-FEB-1992; 92US-0831967.

PA (CETU) CETUS ONCOLOGY CORP.
(CREA-) CREATIVE BIOMOLECULES INC.

PI Houston LL, Huston JS, Oppermann H, Ring DB;

DR WPI; 1993-272889/34.

DR N-PSDB; AAQ46088.

PT New single chain Fv polypeptide binding to C-erbB-2 tumour
antigen - for imaging or treating breast or ovarian cancer etc.

PS Example; pages 71-74; 87pp; English.

CC C-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). A single chain Fv (SFV)
CC is a covalently linked VH-VL heterodimer which is expressed from
CC a gene fusion including VH- and VL- encoding genes connected by
CC a peptide-encoding linker. Such linker sequences are set forth in
CC AA residues 116-135 in AAR39569, which includes part of the 16 AA
CC linker sequences in AAR39572. Using AAQ46084 for the 520C9 monoclonal
CC antibody, a single chain polypeptide can be produced having a
CC binding affinity for a c-erbB-2 related antigen. In another design,
CC the pseudomonal exotoxin fragment analogous to ricin A chain, PE40,
CC is fused to the carboxy terminus of the antic-C-erbB-2 741 SFV.
CC the resulting 7418 SFV-PE40 is a single-chain Fv-toxin fusion
CC protein, which was constructed with an 18 residue short FB leader
CC which initially was left on the protein. 'X' in AAR39573 refers to
CC posn. of stop codon in cDNA.

SO Sequence 667 AA;

Query Match 92.0%; Score 46; DB 14; Length 667;

Best Local Similarity 90.9%; Pred. No. 2;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTAVA 11
| | | | | | | | | |
DB 159 kasqdvstava 169

RESULT 9

AAAR74946
ID AAR74946 standard; peptide; 11 AA.

AC AAR74946;

DT 19-JAN-1996 (first entry)

DE L-CDR-1 of anti-idiotype antibody against human anticancer antibody.

KW Antibody; cancer; CDR; heavy chain; light chain; immunoglobulin;
complementarity determining region.

OS Mus sp.

PN JP07101999-A.

PD 18-APR-1995.

PF 06-OCT-1993; 93JP-0272950.

PR 06-OCT-1993; 93JP-0272950.

PA (HAGI/) HAGIWARA Y.

DR WPI; 1995-182987/24.

PT Novel anti-idiotype antibody against an human anticancer monoclonal
antibody - and DNA sequences encoding the antibody, useful in
pharmacology, medicine and biochemical fields.

PS Claim 11; Page 4; 28pp; Japanese.

CC A new anti-idiotype antibody against a human anticancer monoclonal
CC antibody is claimed. This antibody contains in its heavy chain 3
CC complementarity determining regions CDR1 (AAR74929-R74931), CDR2
CC (AAR74932-R74935) and CDR3 (AAR74936-R74939). This is also true of the
CC light chain which has its own CDR1 (AAR74944-R74946 and AAR85774), CDR2
CC (AAR74947-R74949) and CDR3 (AAR74950-R74954). The antibody and DNA
CC encoding it are useful in pharmacological, medical and biochemical
CC fields.

SO Sequence 11 AA;

Query Match 90.0%; Score 45; DB 16; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.047;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KASQDVTAVA 11
| | | | | | | | | |
DB 1 kasqdvstava 11

RESULT 10

AAAW62189
ID AAW62189 standard; peptide; 11 AA.

AC AAW62189;

DT 21-SEP-1998 (first entry)

DE Mouse anti-HM1.24 antibody L chain V region CDR 1.

KW Mouse; human; humanised; anti-HM1.24 antibody; myeloma; FR; CDR;

KW framework region; complementarity determining region; antigenicity.
 XX Mus sp.
 XX WO9814580-A1.
 PN 09-APR-1998.
 XX 03-OCT-1997; 97WO-JP03553.
 XX 04-OCT-1996; 96JP-0264756.
 XX (CHUS) CHUGAI SEIYAKU KK.
 PA Koishihara Y, Kosaka M, Ohtomo T, Ono K, Tsuchiya M;
 PI Yoshimura Y;
 XX WPI: 1998-286421/25.
 DR Humanised anti-HM1.24 antibody - for treatment of myeloma
 PT Claim 10; Page 102; 210pp; Japanese.
 PS A humanised anti-HM1.24 antibody has been developed which comprises
 XX human L and H chain C regions, and L and/or H chain V regions
 CC containing material originating in mouse anti-HM1.24 antibody. The V
 CC regions contain framework (FR) regions of human origin and
 CC complementarity determining regions (CDR) of mouse origin, leading to
 CC a reshaped humanised antibody. The C regions are human Ck (L-chain) and
 CC human C gamma (especially C gamma 1) (H-chain). The FR regions of the
 CC L chain V region are derived from human subtype HSG1 (e.g. from human
 CC antibody RE1) and the FR regions of the H chain V region are derived
 CC from human subtype HSG1 (e.g. FR1-3 from human antibody HG3 and FR4
 CC from human antibody JH6). The present sequence represents mouse L chain
 CC V region CDR 1 from the present invention. The antibodies are used for
 CC the treatment of myeloma, especially by injection, intravenously,
 CC intramuscularly or subcutaneously. The antibodies are used at 0.01-1000
 CC (especially 5-100) mg/kg body weight. The humanised antibody has low
 CC antigenicity and is therefore effective therapeutically in humans.
 CC
 XX Sequence 11 AA:
 SQ
 QY 1 KASQDVTTAVA 11
 DB 1 kasqdvntava 11
 Query Match 90.0%; Score 45; DB 19; Length 11;
 Best Local Similarity 90.9%; Pred No. 0.047; 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 RESULT 11
 AAM57577
 ID AAM57577 standard; peptide; 11 AA.
 XX
 AC AAM57577;
 XX
 DT 03-SEP-1998 (first entry)
 XX
 DE Chimeric L chain V region CDR-1 for an antibody against hPTRP.
 XX
 KW Chimeric; antibody; human parathormone related peptide; hPTRP; mouse;
 KW L chain; H chain; hypercalcaemia; cancer; malignant lymphoma; CDR;
 KW hypophosphataemia; pathogen; vitamin D resistance; V region; C region;
 KW humanised.
 XX
 OS Synthetic.
 OS Chimeric - Mus sp.
 OS Chimeric - Homo sapiens.
 XX
 PN WO9813388-A1.
 XX

PD 02-APR-1998.
 XX
 XX 24-SEP-1997; 97WO-JP03382.
 XX
 PR 24-JUL-1997; 97JP-0214168.
 PR 26-SEP-1996; 96JP-0255196.
 XX
 PA (CHUS) CHUGAI SEIYAKU KK.
 PI Sato K, Wakahara Y, Yabuta N;
 XX WPI: 1998-230640/20.
 DR
 XX
 PT New chimeric antibodies against human parathormone related
 PR peptide(s) - useful for, e.g. treatment of hypercalcaemia and other
 PT disorders caused by malignant neoplasm(s)
 XX
 PS Claim 9; Page 123; 182pp; Japanese.
 XX
 CC New antibodies have been developed which are specific for human
 CC parathormone related peptides (hPTRP). The antibodies comprise chimeric
 CC L and/or H chains, where the C region is of human and L region of mouse,
 CC origin. The present sequence represents a specifically claimed region of
 CC an antibody of the invention. Host cells, transformed with vectors
 CC containing DNA encoding antibodies of the invention, can be used to
 CC produce the antibodies. The antibodies may be used to treat
 CC hypercalcaemia, especially that due to a malignancy, e.g. cancers of
 CC pancreas, lung, throat, larynx, tongue, gum, oesophagus, stomach, liver,
 CC breast, kidney, bladder, womb or prostate or malignant lymphoma. They
 CC may also be used for treatment of hypophosphataemia such as that due to
 CC pathogens or to vitamin D resistance.
 XX
 SQ Sequence 11 AA:
 QY 1 KASQDVTTAVA 11
 DB 1 kasqdvntava 11
 Query Match 90.0%; Score 45; DB 19; Length 11;
 Best Local Similarity 90.9%; Pred No. 0.047; 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 RESULT 12
 AA02547
 ID AA02547 standard; peptide; 11 AA.
 XX
 AC AA02547;
 XX
 DT 16-JUL-1999 (first entry)
 XX
 DE Artificial CDR(1) of L chain V region of antiHM1.24 antibody.
 XX
 KW Reconstituted human antibody; peptide antigen HM1.24; framework region;
 KW complementary determining region; CDR; anti-HM1.24 antibody; myeloma.
 XX
 OS Synthetic.
 OS
 PN WO9918212-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 02-OCT-1998; 98WO-JP04469.
 XX
 PR 03-OCT-1997; 97JP-0271726.
 XX
 PA (CHUS) CHUGAI SEIYAKU KK.
 PI Tsuchiya M;
 XX WPI: 1999-277273/23.
 DR
 XX

PT Reconstituted human antibody useful in the treatment of myeloma
 XX
 PS Disclosure; Page 82; 256pp; Japanese.
 XX

CC The specification describes a reconstituted human antibody recognizing
 CC the peptide antigen HM1.24. This human antibody contains natural human
 CC framework regions modified by amino acid substitutions to provide
 CC homogeneity with a previously designed framework region (which may
 CC arise from a human or non-human source); and complementary determining
 CC regions (CDR) derived from a non-human anti-HM1.24 antibody. The
 CC reconstituted antibody is useful in the treatment of diseases in which
 CC the surface antigen HM1.24 is implicated such as myeloma. The present
 CC sequence is used in the creation of the antibodies of the invention.
 XX

SO Sequence 11 AA;

QY 1 KASQDVTAVA 11
 Db 1 kasqdvntava 11

Query Match 90.0%; Score 45; DB 20; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.047;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 13
 ID AAW89629 standard; peptide; 11 AA.
 AC AAW89629;
 XX
 DT 14-APR-1999 (first entry)
 XX
 DE Mouse humanised antibody #23-57-137-1 light chain CDS1 peptide.
 XX
 KW Human; parathyroid hormone related protein; PTHrP; cachexia; cancer;
 KW Inhibitor; humanised.
 XX
 OS Mus sp.
 OS Synthetic.
 XX
 PN WO951329-A1.
 XX
 PD 19-NOV-1998.
 XX
 PF 13-MAY-1998; 98WO-JP02116.
 XX
 PR 18-JUL-1997; 97JP-0194445.
 PR 15-MAY-1997; 97JP-0125505.
 XX
 PA (CHUS) CHUGAI SEIYAKU KK.
 XX
 PI Ishii K, Sato K, Tunesari T;
 XX
 DR WPI; 1999-070101/06.
 XX
 PT Inhibitors of binding of parathyroid hormone related peptide to its
 PT receptor - useful for e.g. treatment of cachexia arising from
 PT cancer or other diseases
 XX
 PS Example 2; Page 84; 125pp; Japanese.
 XX

CC The present invention describes compositions for the treatment of
 CC cachexia containing a substance which inhibits the binding of a
 CC parathyroid hormone related peptide (PTHrP) to its receptor, as an
 CC active component. This substance may be an antagonist to the receptor,
 CC or an antibody (preferably monoclonal) or antibody fragment.
 CC recognising PTHrP. The antibody is preferably humanised or chimeric.
 CC The present invention also describes a humanised antibody prepared
 CC by hybridoma 23-57-137-1 (FERM BP-5631). The composition is used for
 CC the treatment of cachexia arising in connection with diseases such as
 CC cancer, thereby improving the quality of life of the patient. The

CC present sequence represents mouse humanised antibody light chain CDS1
 CC from #23-57-137-1 from the present invention.
 XX
 SO Sequence 11 AA;

QY 1 KASQDVTAVA 11
 Db 1 kasqdvntava 11

Query Match 90.0%; Score 45; DB 20; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.047;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 14
 ID AAY77515 standard; peptide; 11 AA.
 AC AAY77515;
 XX
 DT 26-APR-2000 (first entry)
 XX
 DE Antibody H chain V region CDRI peptide seq ID No: 59.
 XX
 KW Hypercalcaemic crisis; parathyroid hormone related peptide; PTHrP; tumour.
 KW Homo sapiens.
 XX
 OS Homo sapiens.
 XX
 PN WO200000219-A1.
 XX
 PD 06-JAN-2000.
 XX
 PF 25-JUN-1999; 99WO-JP03433.
 XX
 PR 26-JUN-1998; 98JP-0180143.
 XX
 PA (CHUS) CHUGAI SEIYAKU KK.
 XX
 PI Sato K, Tsunenari T;
 XX
 DR WPI; 2000-117115/10.
 XX
 PT Treatment of hypercalcaemic crisis with a substance inhibiting binding
 PT of parathyroid hormone related peptide to its receptor -
 XX
 PS Example 2; Page 99; 120pp; Japanese.
 XX

CC The invention relates to a method of treatment of hypercalcaemic crisis.
 CC A composition for the treatment of hypercalcaemic crisis contains as
 CC active component a substance which inhibits the binding of parathyroid
 CC hormone related peptide (PTHrP) to its receptor. The inhibitor is used
 CC for the treatment of hypercalcaemic crisis, such as that associated with
 CC a malignant tumour.
 XX

SO Sequence 11 AA;

QY 1 KASQDVTAVA 11
 Db 1 kasqdvntava 11

Query Match 90.0%; Score 45; DB 21; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.047;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 15
 ID AAG67113 standard; peptide; 11 AA.
 AC AAG67113;
 XX
 DT 26-APR-2000 (first entry)
 XX
 DE Antibody H chain V region CDRI peptide seq ID No: 59.
 XX
 KW Hypercalcaemic crisis; parathyroid hormone related peptide; PTHrP; tumour.
 KW Homo sapiens.
 XX
 OS Homo sapiens.
 XX
 PN WO200000219-A1.
 XX
 PD 06-JAN-2000.
 XX
 PF 25-JUN-1999; 99WO-JP03433.
 XX
 PR 26-JUN-1998; 98JP-0180143.
 XX
 PA (CHUS) CHUGAI SEIYAKU KK.
 XX
 PI Sato K, Tsunenari T;
 XX
 DR WPI; 2000-117115/10.
 XX
 PT Treatment of hypercalcaemic crisis with a substance inhibiting binding
 PT of parathyroid hormone related peptide to its receptor -
 XX
 PS Example 2; Page 99; 120pp; Japanese.
 XX

CC The invention relates to a method of treatment of hypercalcaemic crisis.
 CC A composition for the treatment of hypercalcaemic crisis contains as
 CC active component a substance which inhibits the binding of parathyroid
 CC hormone related peptide (PTHrP) to its receptor. The inhibitor is used
 CC for the treatment of hypercalcaemic crisis, such as that associated with
 CC a malignant tumour.
 XX

DT 13-NOV-2001 (first entry)
 DE Amino acid sequence of a human peptide.
 XX
 KW Tissue decomposition inhibitor; parathyroid hormone; cancer cachexia;
 KW septicemia; injury; muscular dystrophy; cytokine; interleukin-6;
 KW granulocyte colony stimulating factor; interleukin-11;
 KW leukemia inhibitory factor; weight loss.
 XX
 OS Homo sapiens.
 XX
 PN WO200164249-A1.
 XX
 PD 07-SEP-2001.
 XX
 PF 30-AUG-2000; 2000WO-JP05886.
 XX
 PR 28-FEB-2000; 2000JP-0052414.
 XX
 PA (CHUS) CHUGAI SEIYAKU KK.
 XX
 PI Saito H, Tsunenari T, Onuma E, Sato K;
 XX
 DR WPI; 2001-550131/61.
 XX
 PT Tissue decomposition inhibitor that prevents parathyroid hormone
 PT associated proteins from binding to its receptor
 XX
 PS Disclosure: Page 108; 132pp; Japanese.
 XX
 CC The specification describes a tissue decomposition inhibitor, which
 CC comprises a substance that inhibits peptides associated with
 CC parathyroid hormone (PTH) from binding with their receptor. The method
 CC is used to inhibit tissue decomposition caused by cancer cachexia,
 CC septicemia, heavy external injury or muscular dystrophy, and for
 CC treating patients with elevated cytokine (interleukin-6, granulocyte
 CC colony stimulating factor, interleukin-11 and leukemia inhibitory
 CC factor) levels. It may also be used for preventing weight loss caused
 CC by cancer cachexia. The present sequence represents a peptide, which is
 CC used in the course of the invention.
 CC
 SQ Sequence 11 AA:
 QY 1 KASQDVTTAVA 11
 DB 1 kasqdvttava 11

Query Match 90.0%; Score 45; DB 22; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.047;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Search completed: August 20, 2002, 11:34:38
 Job time: 632 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:28:41 : Search time 25.74 seconds
(Without alignments)
41.064 Million cell updates/sec

Title: US-09-824-286-4_COPY_24_34
Perfect score: 50
Sequence: 1 KASQDVTAVA 11

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	94.0	152	2 S30751	Ig kappa chain pre
2	46	92.0	93	2 S38559	Ig light chain V r
3	46	92.0	107	2 S32191	Ig kappa chain V r
4	45	92.0	107	2 S32192	Ig kappa chain V r
5	45	90.0	107	2 S09967	Ig kappa chain V-J
6	43	86.0	117	2 S42466	Ig kappa chain V r
7	42	84.0	119	2 P00265	Ig kappa chain V r
8	42	84.0	149	1 KYMS11	Ig kappa chain pre
9	41	82.0	106	2 PL0088	Ig kappa chain V r
10	40	80.0	98	2 PH1073	Ig light chain V r
11	39	78.0	94	2 F33730	Ig light chain V r
12	38	76.0	82	2 S38560	Ig light chain V r
13	38	76.0	90	2 S38561	Ig light chain V r
14	38	76.0	108	2 PL0083	Ig kappa chain V r
15	38	76.0	131	2 PL0207	anti-idiotypic ant
16	37	74.0	100	2 H38601	Ig kappa chain V r
17	36	72.0	1449	2 T30857	glucosyltransferas
18	36	72.0	1449	2 T30552	glucosyltransferas
19	34	68.0	45	2 B43859	ORF overlapping ca
20	34	68.0	108	2 PL0204	anti-DNA autoantib
21	34	68.0	328	2 AB1781	hypothetical prote
22	34	68.0	328	2 AC1405	hypothetical prote
23	34	68.0	374	2 AC0469	hypothetical prote
24	34	68.0	389	2 AF0920	probable uroporphyr
25	34	68.0	393	2 S02185	uroporphyrin-III C
26	34	68.0	395	2 T21679	hypothetical prote
27	34	68.0	399	2 E91220	uroporphyrinogen I
28	34	68.0	399	2 G86066	uroporphyrinogen I
29	34	68.0	435	2 H87516	D-alanyl-D-alanine

30	34	68.0	501	2 F98050	lysozyme (EC 3.2.1
31	34	68.0	633	2 E31265	lipa protein - Shi
32	34	68.0	815	2 T00546	serine/threonine-s
33	33	66.0	100	2 T50611	hypothetical prote
34	33	66.0	226	2 T29210	hypothetical prote
35	33	66.0	485	1 B45343	glycoprotein gp13
36	33	66.0	485	2 T42559	envelope protein 1
37	33	66.0	822	2 T25866	hypothetical prote
38	33	66.0	2396	2 T13714	kakapo gene protei
39	32	64.0	165	2 JC5737	membrane-bound pro
40	32	64.0	191	2 AG2916	conserved hypotet
41	32	64.0	235	2 B97691	pHr protein (Af02
42	32	64.0	263	2 D83085	conserved hypotet
43	32	64.0	330	2 T41917	hypothetical prote
44	32	64.0	371	2 E84709	hypothetical prote
45	32	64.0	398	2 T26274	hypothetical prote

ALIGNMENTS

RESULT 1
S30751
Ig kappa chain precursor V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Jan-2000
C:Accession: S30751
R:Grant, F.J.; Levin, S.D.; Gilbert, T.; Kindsvogel, W.
Nucleic Acids Res. 15, 5496, 1987
A:Title: Improved RNA sequencing method to determine immunoglobulin mRNA sequence.
A:Reference number: S30751; MUID:87260030
A:Accession: S30751
A:Molecule type: mRNA
A:Residues: 1-152 <GBR>
A:Cross-references: EMBL:X05877; NID:952195; PIDN:CAA29301.1; PID:952196
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: heterotetramer; Immunoglobulin
F:45-119/Domain: Immunoglobulin homology <IMM>
F:138-152/Domain: C region (C-kappa) (fragment) #status predicted <CRE>

Query Match 94.0%; Score 47; DB 2; Length 152;
Best Local Similarity 90.9%; Pred. No. 0.042;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KASQDVTAVA 11
DB 53 KASQDVTAVS 63
RESULT 2
S38559
Ig light chain V region (ASWAI) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
R:Monestier, M.; Losman, L.J.; Novick, K.E.; Aris, J.P.
Submitted to the EMBL Data Library, September 1993
A:Description: Molecular analysis of mercury-induced anti-nuclear antibodies in R-
A:Reference number: S38559
A:Accession: S38559
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-93 <MON>
A:Cross-references: EMBL:X75101; NID:9414145; PIDN:CAA52992.1; PID:9414146
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: immunoglobulin
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 92.0%; Score 46; DB 2; Length 93;
Best Local Similarity 90.9%; Pred. No. 0.04;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11
|||||:||||
Db 24 KASQDVSTAVA 34

RESULT 3

S32191
Ig kappa chain V region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 21-Jan-2000
C:Accession: S32191
R:Izui, S.
Submitted to the EMBL Data Library, February 1993
A:Reference number: S32185
A:Accession: S32191
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-107 <I2U>
A:Cross-references: EMBL:X70095; NID:9288260; PIDN:CAA49700.1; PID:9288261
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 92.0%; Score 46; DB 2; Length 107;
Best Local Similarity 90.9%; Pred. No. 0.047;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11
|||||:||||
Db 24 KASQDVSTAVA 34

RESULT 4

S32192
Ig kappa chain V region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 21-Jan-2000
C:Accession: S32192
R:Izui, S.
Submitted to the EMBL Data Library, February 1993
A:Reference number: S32185
A:Accession: S32192
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-107 <I2U>
A:Cross-references: EMBL:X70097; NID:9288262; PIDN:CAA49701.1; PID:9288263
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 92.0%; Score 46; DB 2; Length 107;
Best Local Similarity 90.9%; Pred. No. 0.047;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11
|||||:||||
Db 24 KASQDVSTAVA 34

RESULT 5

S09967
Ig kappa chain V-J region (31-9D) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 16-Aug-1996
C:Accession: S09967
R:Reininger, L.; Shibata, T.; Ozaki, S.; Shirai, T.; Jaton, J.C.; Izui, S.
Eur. J. Immunol. 20, 771-777, 1990
A:Title: Variable region sequences of pathogenic anti-mouse red blood cell autoantibodies
A:Reference number: S09955; MUID:9026328
A:Accession: S09967

A:Molecule type: mRNA
A:Residues: 1-107 <RET>
A:Cross-references: EMBL:X51855
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin

Query Match 90.0%; Score 45; DB 2; Length 107;
Best Local Similarity 90.9%; Pred. No. 0.074;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11
|||||:||||
Db 24 KASQDVNTAVA 34

RESULT 6

S42466
Ig kappa chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
C:Accession: S42466
R:Shyanov, P.A.; Bessalov, I.A.; Terletskaya, H.N.; Deyev, S.M.
Submitted to the EMBL Data Library, March 1994
A:Reference number: S42466
A:Accession: S42466
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-117 <SHI>
A:Cross-references: EMBL:X78108; NID:9460824; PIDN:CAA54998.1; PID:9460825
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:26-100/Domain: immunoglobulin homology <IMM>

Query Match 86.0%; Score 43; DB 2; Length 117;
Best Local Similarity 90.9%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11
|||||:||||
Db 34 KASQDVGTAVA 44

RESULT 7

PQ0265
Ig kappa chain V region (MC1) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 21-Jan-2000
C:Accession: PQ0265
R:Lothman, K.L.; Carrillo, M.A.; Kennedy, R.C.
Gene 105, 283-284, 1991
A:Title: Sequence analysis of the variable region of a mouse gene encoding a monoclonal antibody
A:Reference number: PQ0265; MUID:92039046
A:Accession: PQ0265
A:Molecule type: mRNA
A:Residues: 1-119 <LOH>
A:Cross-references: GB:M59985
C:Comment: This protein recognizes a restricted idiotype associated with antibodies s
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:28-102/Domain: immunoglobulin homology <IMM>
F:36-46/Region: complementarity-determining 1
F:62-68/Region: complementarity-determining 2
F:101-109/Region: complementarity-determining 3

Query Match 84.0%; Score 42; DB 2; Length 119;
Best Local Similarity 81.8%; Pred. No. 0.34;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11
|||||:||||

Db 36 KASODVSTTVA 46

RESULT 8

KVMS11

Ig kappa chain precursor V region (MPC11) - mouse

C:Species: Mus musculus (house mouse)

C:Date: 19-Feb-1984 #sequence_revision 19-Feb-1984 #text_change 21-Jan-2000

C:Accession: A90823; A90753; A90298; A01916

R:Kelley, D.E.; Coleclough, C.; Perry, R.P.

Cell 29, 681-689, 1982

A:Title: Functional significance and evolutionary development of the 5'-terminal regions

A:Reference number: A90823; MWID:83001944

A:Accession: A90823

A:Molecule type: DNA

A:Residues: 1-71 <KRL>

A>Note: the sequence was determined from the differentiated gene

R:Rabbits, T.H.; Hamlyn, P.H.; Mathysens, G.; Roe, B.A.

Can. J. Biochem. 58, 176-187, 1980

A:Title: The variability, arrangement, and rearrangement of immunoglobulin genes.

A:Reference number: A90753; MWID:80176554

A:Accession: A90753

A:Molecule type: mRNA

A:Residues: 41-149 <RAB>

R:Smith, G.P.

Biochem. J. 171, 337-347, 1978

A:Title: Sequence of the full-length immunoglobulin kappa-chain of mouse myeloma MPC 11.

A:Reference number: A90298; MWID:78186617

A:Accession: A90298

A:Contents: myeloma protein MPC11

A:Molecule type: protein

A:Residues: 30-149 <SMT>

A>Note: the amidation states of residues 58, 69, 101, 111, 120, 122, and 123 were not de

C:Comment: The mature chain has 12 additional residues at its amino end, due to a tandem

42 corresponds to the amino-terminal residue of typical kappa chains.

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kapp

hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: duplication; heterotetramer; immunoglobulin

F:1-29/Domain: signal sequence #status predicted <Sig>

F:30-149/Product: Ig kappa chain V region (MPC11) #status experimental <MAT>

F:57-131/Domain: immunoglobulin homology <IMM>

F:64-129/Disulfide bonds: #status predicted

Query Match

Best Local Similarity

Matches 9; Conservative

1; Mismatches

1; Indels

0; Gaps

0;

OY

1 KASODVTTTAVA 11

Db 65 KASODVSTTVA 75

RESULT 9

PL0088

Ig kappa chain V regions (12S18-1, 12S28-16) - mouse

C:Species: Mus musculus (house mouse)

C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 21-Jan-2000

C:Accession: PL0088

R:Meek, K.; Hasemann, C.; Pollok, B.; Alkan, S.S.; Bralt, M.; Slaoui, M.; Urbain, J.; Ca

J. Exp. Med. 169, 519-533, 1989

A:Title: Structural characterization of antidiotytic antibodies; evidence that Ab2s are

A:Reference number: PL0080; MWID:89094248

A:Accession: PL0088

A:Molecule type: mRNA

A:Residues: 1-106 <MEE>

A:Cross-references: GB:X58581; GB:Y00794; NID:951592; PIDN:CAA41457.1; PID:9330151

A>Note: the sequence shown here is from the V kappa regions an antidiotytic monoclonal

C:Superfamily: immunoglobulin V region; immunoglobulin homology

Query Match

Best Local Similarity

Matches 9; Conservative

1; Mismatches

1; Indels

0; Gaps

0;

OY

1 KASODVTTTAVA 11

Db 24 KASODVNSAVA 34

RESULT 10

PH1073

Ig light chain V region (clone 17s.130) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jan-2000

C:Accession: PH1073

R:Fillman, D.M.; Jou, N.T.; Hill, R.J.; Marlon, T.N.

J. Exp. Med. 176, 761-779, 1992

A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective

A:Reference number: PH0971; MWID:92381444

A:Accession: PH1073

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-98 <TIL>

A:Experimental source: B cell, strain [NZB x NZW]F1

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: immunoglobulin

F:16-90/Domain: immunoglobulin homology <IMM>

Query Match

Best Local Similarity

Matches 9; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

OY

1 KASODVTTTAVA 11

Db 24 KASODVNAVA 34

RESULT 11

F33730

Ig kappa chain V region (19.34) - mouse

C:Species: Mus musculus (house mouse)

C:Date: 09-Mar-1990 #sequence_revision 18-Sep-1992 #text_change 21-Jan-2000

C:Accession: F33730

R:Lawler, A.M.; Kearney, J.F.; Kuehl, M.; Gearhart, P.J.

Proc. Natl. Acad. Sci. U.S.A. 86, 6744-6747, 1989

A:Title: Early rearrangements of genes encoding murine immunoglobulin kappa-chains, "

A:Reference number: A33730; MWID:89367325

A:Accession: F33730

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-94 <LAW>

A:Cross-references: GB:M26002; NID:9197123; PIDN:AAA38918.1; PID:9554119

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:16-90/Domain: immunoglobulin homology <IMM>

Query Match

Best Local Similarity

Matches 9; Conservative

1; Mismatches

1; Indels

0; Gaps

0;

OY

1 KASODVTTTAVA 11

Db 24 KASODVTTTAVA 34

RESULT 12

S38560

Ig light chain V region (ASWB1) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 23-Jul-1999

C:Accession: S38560
R:Monestier, M.; Losman, L.J.; Novick, R.E.; Aris, J.P.
Submitted to the EMBL Data Library, September 1993
A:Description: Molecular analysis of mercury-induced anti-nuclear antibodies in H-2s M
A:Reference number: S38559
A:Accession: S38560
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-82 <MON>
A:Cross-references: EMBL:X75102; NID:9414147; PIDN:CAA52993.1; PID:9414148
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: immunoglobulin

Query Match 76.0%; Score 38; DB 2; Length 82;
Best Local Similarity 81.8%; Pred. No. 1.5;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KASQDVTTAVA 11
DB 24 KASQNVGTAVA 34

RESULT 13
S38561
Ig light chain V region (ASMP1) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
C:Accession: S38561
R:Monestier, M.; Losman, L.J.; Novick, R.E.; Aris, J.P.
Submitted to the EMBL Data Library, September 1993
A:Description: Molecular analysis of mercury-induced anti-nuclear antibodies in H-2s M
A:Reference number: S38559
A:Accession: S38561
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-90 <MON>
A:Cross-references: EMBL:X75103; NID:9414155; PIDN:CAA52994.1; PID:9414156
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: immunoglobulin
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 76.0%; Score 38; DB 2; Length 90;
Best Local Similarity 81.8%; Pred. No. 1.6;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KASQDVTTAVA 11
DB 24 KASQNVGTAVA 34

RESULT 14
PL0083
Ig kappa chain V region (E3) - mouse
C:Species: Mus musculus (house mouse)
C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 21-Jan-2000
C:Accession: PL0083
R:Meek, K.; Hasemann, C.; Pollok, B.; Alkan, S.S.; Bralt, M.; Slaoui, M.; Urbain, J.; Ca
J. Exp. Med. 169, 519-533, 1989
A>Title: Structural characterization of antidiotytic antibodies: evidence that Ab2s are
A:Reference number: PL0080; MUID:89094248
A:Accession: PL0083
A:Molecule type: mRNA
A:Residues: 1-108 <MEK>
A:Experimental source: strain BALB/c
A>Note: the sequence shown here is from the V kappa region of an antidiotytic monoclonal
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:16-91/Domain: immunoglobulin homology <IMM>

Query Match 76.0%; Score 38; DB 2; Length 108;

Best Local Similarity 81.8%; Pred. No. 2;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KASQDVTTAVA 11
DB 24 KASQNVGTAVA 34

RESULT 15
PL0207
anti-idiotypic antibody E225, kappa chain V region precursor - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 21-Jan-2000
C:Accession: PL0207
R:Souchon, H.; Doyen, N.; Riottot, M.M.; Rougeon, F.; Poljak, R.J.
Mol. Immunol. 27, 429-433, 1990
A>Title: Nucleotide sequence of the VH, VL regions of an anti-idiotypic antibody reac
A:Reference number: PL0207; MUID:90309764
A:Accession: PL0207
A:Molecule type: mRNA
A:Residues: 1-131 <SGU>
A>Note: This sequence corresponds to subgroup V mouse immunoglobulin light chain
C:Superfamily: immunoglobulin V region; immunoglobulin homology
F:1-20/Domain: signal sequence #status predicted <SIG>
F:36-110/Domain: immunoglobulin homology <IMM>
F:44-54/Region: complementarity-determining 1
F:70-76/Region: complementarity-determining 2
F:109-117/Region: complementarity-determining 3
F:116-127/Region: J region

Query Match 76.0%; Score 38; DB 2; Length 131;
Best Local Similarity 81.8%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 KASQDVTTAVA 11
DB 44 KASQDVTAVA 54

Search completed: August 20, 2002, 11:35:17
Job time: 396 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:34:41 ; Search time 15.95 Seconds

(Without alignments)
26.703 Million cell updates/sec

Title: US-09-824-286-4_COPY_24_34
Perfect score: 50
Sequence: 1 KASODVTAVA 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	84.0	149	KV5A_MOUSE	P01633 mus musculu
2	34	68.0	139	HEMX_ECOLI	P09127 escherichia
3	34	68.0	633	IPAA_SHIFL	P18010 shigella fl
4	33	66.0	485	VGIC_HSYE4	P22596 equine herp
5	33	66.0	975	CSEL_DROME	P09201 drosophila
6	32	64.0	165	ATPF_STROM	P35785 streptococc
7	32	64.0	474	FLA_LEGMI	P33606 legionella
8	32	64.0	481	HTR3_HALSA	P09400 halobacteri
9	32	64.0	108	KV1Y_HUMAN	P80362 homo sapien
10	31	62.0	109	GV12_HAIAN	P09400 halobacteri
11	31	62.0	279	TCPA_CAEEL	P41988 caenorhabdi
12	31	62.0	549	TCPA_CAEEL	P41988 caenorhabdi
13	31	62.0	788	ACON_CAEEL	P44455 caenorhabdi
14	31	62.0	1041	EGT2_YEAST	P28235 saccharomyc
15	31	62.0	1275	YAU9_SCHPO	P48946 synechocyst
16	30	60.0	71	RS18_SYNY3	P00607 candida tro
17	30	60.0	160	VATL_CANTR	P41847 caenorhabdi
18	30	60.0	238	Y097_CAEEL	P39039 mus musculu
19	30	60.0	239	MABA_MOUSE	P24076 medicago sa
20	30	60.0	325	GBLP_MEDSA	P03936 glycyne max
21	30	60.0	325	GBLP_SOYBN	P06762 enterococcu
22	30	60.0	442	FTSA_ENTHR	P08369 escherichia
23	30	60.0	450	CRPD_ECOLI	P09127 escherichia
24	30	60.0	459	PEDA_LACLA	P12889 equine herp
25	30	60.0	468	VGIC_HSYE4	P22596 equine herp
26	30	60.0	483	NDAD_ALICX	P35887 caulobacter
27	30	60.0	490	DNAA_CAUCR	P42122 alcaligenes
28	30	60.0	498	NDAD_ALICX	P35887 caulobacter
29	30	60.0	562	CHIT_YEAST	P29029 saccharomyc
30	30	60.0	750	LEU2_YEAST	P55811 rhyzopus ni
31	30	60.0	793	NETB_DROME	P24568 drosophila
32	30	60.0	901	VGLB_GPCMV	P05902 guinea pig
33	30	60.0	969	AGLU_SCHPO	P09004 schizosacch

34	30	60.0	984	1	SV13_MOUSE	Q04891 mus musculu
35	30	60.0	2334	1	WAPA_BACSU	Q07833 bacillus su
36	30	60.0	2512	1	POLN_SINDV	P03317 sindbis vir
37	30	60.0	2514	30	POLN_SINDV	P27283 sindbis vir
38	29	58.0	81	1	Y039_STNP7	P27368 synechococc
39	29	58.0	107	1	FER_PSLA	P24806 psalteriomo
40	29	58.0	145	1	LIRP_LOCOM	P15131 locusta mig
41	29	58.0	181	1	RBS1_BRANA	P05346 brassica na
42	29	58.0	204	1	PS21_ARATH	Q23714 arabidopsis
43	29	58.0	222	1	SODM_MOUSE	P06671 mus musculu
44	29	58.0	222	1	SODM_RAT	P07895 rattus norv
45	29	58.0	247	1	I43B_SOYBN	Q96451 glycine max

ALIGNMENTS

RESULT	ID	Sequence	Standard	PRT	AA
1	KV5A_MOUSE	21-JUL-1986 (Rel. 01, Created)			
2	P01633	21-JUL-1986 (Rel. 01, Last sequence update)			
3	DT	15-JUL-1999 (Rel. 38, Last annotation update)			
4	DE	Ig kappa chain V-V region MPC11 precursor.			
5	OS	Mus musculus (Mouse).			
6	OC	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:			
7	OC	Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Mus.			
8	OX	NCBI_Taxid=10090;			
9	RN	[1]			
10	RP	SEQUENCE OF 1-71 FROM N.A.			
11	RX	MEDLINE=83001944; PubMed=6288267;			
12	RA	Kelley D.E., Coleclough C., Perry R.P.;			
13	RT	"Functional significance and evolutionary development of the			
14	RT	5'-terminal regions of immunoglobulin variable-region genes."			
15	RL	Cell 29:681-689(1982).			
16	RN	[2]			
17	RP	SEQUENCE OF 41-149 FROM N.A.			
18	RX	MEDLINE=80176554; PubMed=6245773;			
19	RA	Rabbits T.H., Hamlyn P.H., Matlyssens G., Roe B.A.;			
20	RT	"The variability, arrangement, and rearrangement of immunoglobulin			
21	RT	genes."			
22	RL	Can. J. Biochem. 58:176-187(1980).			
23	RN	[3]			
24	RP	SEQUENCE OF 30-149.			
25	RX	MEDLINE=78186617; PubMed=418775;			
26	RA	Smith G.P.;			
27	RT	"Sequence of the full-length immunoglobulin kappa-chain of mouse			
28	RT	myeloma MPC 11."			
29	RL	Biochem. J. 171:337-347(1978).			
30	CC	MISCELLANEOUS: THE MATURE CHAIN HAS 12 ADDITIONAL RESIDUES AT ITS			
31	CC	AMINO END, DUE TO A TANDER DUPLICATION OF 36 NUCLEOTIDES AFTER THE			
32	CC	CODON FOR RESIDUE 36. RESIDUE 42 CORRESPONDS TO THE AMINO-TERMINAL			
33	CC	RESIDUE OF TYPICAL KAPPA CHAINS.			
34	CC	-----			
35	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
36	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
37	CC	the European Bioinformatics Institute. There are no restrictions on its			
38	CC	use by non-profit institutions as long as its content is in no way			
39	CC	modified and this statement is not removed. Usage by and for commercial			
40	CC	entities requires a license agreement (See http://www.isb.ch/announce/			
41	CC	or send an email to license@isb-sib.ch).			
42	CC	-----			
43	DR	EMBL: J00561, AAA8776.1; -.			
44	DR	PIR: A01916; KVM51.			
45	DR	InterPro: IPR003006; IG_MHC.			
46	DR	InterPro: IPR003596; IG_V.			
47	DR	Pfam: PF00047; IG: 1.			
48	DR	SMART: SM00406; IGV: 1.			
49	DR	Immunoglobulin V region: Signal; Repeat.			
50	FT	CHAIN	1	29	
51	FT	SIGNAL	30	149	IG KAPPA CHAIN V-V REGION MPC11.
52	FT	DOMAIN	42	64	FRAMEWORK-1.

```

FT DOMAIN 65 75 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 76 90 FRAMEWORK-2.
FT DOMAIN 91 97 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 98 129 FRAMEWORK-3.
FT DOMAIN 130 138 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 139 148 FRAMEWORK-4.
FT REPEAT 26 35
FT REPEAT 38 47
FT NON_TER 149 149
SQ SEQUENCE 149 AA; 16434 MW; B0480C87B682AC3E CRC64;

Query Match 84.0%; Score 42; DB 1; Length 149;
Best Local Similarity 81.8%; Pred. No. 0.14;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Cy 1 KASODVTTAVA 11
    |||||:|
Db 65 KASODVSTTVA 75

RESULT 2
HEMX_ECOLI STANDARD; PRT: 393 AA.
ID P09127;
AC 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative uroporphyrin-III C-methyltransferase (EC 2.1.1.107) (Urogen
DE III methylase) (ORF X).
GN HMXX OR B3803.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89098348; PubMed=3062586;
RA Sasarman A., Echelard Y., Letowski J., Tardif D., Drolet M.;
RT "Nucleotide sequence of the hemx gene, the third member of the uro
RT operon of Escherichia coli K12."
RL Nucleic Acids Res. 16:11835-11835(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / CS520;
RX MEDLINE=89041586; PubMed=3054815;
RA Alefounder P.R.;
RT "The sequence of hemc, hemd and two additional E. coli genes."
RL Nucleic Acids Res. 16:9871-9871(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=92358234; PubMed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes."
RL Science 257:771-778(1992).
RN [4]
RP SEQUENCE OF 1-12.
RC STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12."
RL Electrophoresis 18:1259-1313(1997).
CC -1- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + uroporphyrin III =
CC 2 S-adenosyl-L-homocysteine + 5-hydroxytryptophan.
CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF SIROHEME AND COBALAMIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13406; CAA31772.1; -
DR EMBL; X12614; CAA31134.1; -
DR EMBL; M87049; AAA67599.1; -
DR EMBL; AE000456; AAC76806.1; -
DR PIR; S02185; S02185.
DR PIR; S30693; S30693.
DR EC02DBASE; B043.0; 6TH EDITION.
DR EC02DBASE; B043.1; 6TH EDITION.
DR EcGene; EG10433; hemx.
KW Porphyrin biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 393 AA; 42963 MW; 9D272C6401D0E354 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 393;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Cy 2 ASODVTTAVA 11
    |||||:|
Db 151 SDODVTTAA 160

RESULT 3
IPAA_SHIFL STANDARD; PRT: 633 AA.
ID IPAA_SHIFL
AC P18010;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 70 kDa antigen.
DE IPAA.
OS Shigella flexneri.
OC Plasmid 210 kb invasion pWR100.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_Taxid=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90T / SEROTYPE 5;
RX MEDLINE=90221912; PubMed=2183200;
RA Venkatesan M.M., Buysse J.M.;
RT "Nucleotide sequence of invasion plasmid antigen gene ipaA from
RT Shigella flexneri 5."
RL Nucleic Acids Res. 18:1648-1648(1990).
RN [2]
RP SEQUENCE OF 1-88 FROM N.A.
RC STRAIN=M90T / SEROTYPE 5;
RX MEDLINE=89057927; PubMed=3057506;
RA Venkatesan M.M., Buysse J.M.;
RT "Characterization of invasion plasmid antigen genes (ipaBCD) from
RT Shigella flexneri."
RL Proc. Natl. Acad. Sci. U.S.A. 85:9317-9321(1988).
CC -1- FUNCTION: ASSOCIATED WITH THE ENTRY OF THE BACTERIA INTO COLONIC
CC EPITHELIAL CELLS.
CC -----
CC MISCELLANEOUS: SYNTHESIS OF THIS IMMUNOGEN IS RERESSED AT 30
CC DEGREES CELSIUS AND RESTORED AT 37 DEGREES CELSIUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17628; CAA35624.1; -
DR EMBL; J04117; AAA26525.1; -

```

DR PIR: E31265; E31265.
DR PIR: S12763; S12763.
KW Antigen: Plasmid: Virulence.
SQ SEQUENCE 633 AA; 70093 MW; 2F804F45355E4751 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 633;
Best Local Similarity 54.5%; Pred. No. 27;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQVTTAV 11
11:111111:
DB 614 KAAQVTTSL 624

RESULT 4
VGLC_HSV4 STANDARD; PRT; 485 AA.
ID VGLC_HSV4
AC P22596;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 30-MAR-2000 (Rel. 39, Last annotation update)
DE Glycoprotein C precursor (Glycoprotein 13).
GN GC OR GP13.
OS Equine herpesvirus type 4 (strain 1942) (EHV-4) (Equine herpesvirus
OS type 1 subtype 2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=10333;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021040; PubMed=2171212;
RA Nicolson L., Onions D.E.;
RT "The nucleotide sequence of the equine herpesvirus 4 gc gene
RT homologue".
RL Virology 179:378-387(1990).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC
DR EMBL: M58031; AAA46083.1; -
DR EMBL: A21044; CA01528.1; -
DR PIR: B45343; B45343.
DR InterPro: IPR001038; EHV_gp13.
DR InterPro: IPR001654; Marek_A.
DR Pfam: PF02124; Marek_A; 1.
DR PRINTS: PR00668; GlycoproteinC.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 32
FT CHAIN 33 485 GLYCOPROTEIN C.
FT DOMAIN 33 444 EXTRACELLULAR.
FT TRANSMEM 445 468
FT CARBOHYD 60 60
FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 485 AA; 52509 MW; 599AC10EB9C2598E CRC64;

Query Match

66.0%; Score 33; DB 1; Length 485;

Best Local Similarity 87.5%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 SQVTTAV 10
11111111
DB 382 SQVTTGV 389

RESULT 5
CSEL_DROME STANDARD; PRT; 975 AA.
ID CSEL_DROME
AC Q9XZU1; Q9VJH4; Q9UB14;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Importin-alpha re-exporter (Cellular apoptosis susceptibility protein
DE homolog).
DE CAS OR BCDA:ID14270 OR CG13281.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Tekotte H., Davis I.;
RT "Drosophila melanogaster CAS".
RT Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Prankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandal D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostreffi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skuski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster".
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Tsang G., Brokstein P., Friese E., Harvey D., Evans-Holm M.,

RA Lewis S.E., Suh C., Rubin G.M.;
RT "Full Length Drosophila melanogaster cDNA sequence."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Export receptor for importin alpha. Mediates importin-
CC alpha reexport from the nucleus to the cytoplasm after import
CC substrates have been released into the nucleoplasm (By
CC similarity).
CC -1- SUBUNIT: Binds with high affinity to importin-alpha only in the
CC presence of RanGTP (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CSE1 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AJ238857; CAB42967.1; -
DR EMBL: AE003652; AAF53575.1; -
DR EMBL: AF132562; AAD27861.1; -
DR FLYBASE: FBgn0022213; Cas.
DR InterPro: IPR001494; IBN_NT.
KW Transport; Protein transport; Nuclear protein.
FT CONFLICT 194 194 K -> T (IN REF. 2).
FT CONFLICT 331 331 Q -> P (IN REF. 1).
FT CONFLICT 589 590 AA -> GR (IN REF. 1).
SQ SEQUENCE 975 AA; 110168 MW; 5C9BAllC08D49C11 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 975;
Best Local Similarity 63.6%; Pred. No. 69;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 KASQDVTAVV 11
Db 52 KAQMDTRVVA 62
|||:|:|:|

RESULT 6
ATPE_STRMU STANDARD; PRT; 165 AA.
ID ATPE_STRMU
AC P95785;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP synthase B chain (EC 3.6.3.14).
GN ATPF.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=97149283; PubMed=8996091;
RA Smith A.J., Quivey R.G., Faustoferrl R.C.;
RT "Cloning and nucleotide sequence analysis of the Streptococcus mutans
RT membrane-bound, proton-translocating ATPase operon."
RL Gene 183:87-96(1996)
RU
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U31170; AAD13379.1; -
DR InterPro: IPR002146; ATP-synt_B.
DR Pfam: PF00430; ATP-synt_B.1.
KW Hydrogen ion transport; Transmembrane; CF(0).
FT TRANSMEM 7 27 POTENTIAL.
SQ SEQUENCE 165 AA; 17952 MW; BIDAC1206386113 CRC64;

Query Match 64.0%; Score 32; DB 1; Length 165;
Best Local Similarity 54.5%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KASQDVTAVV 11
Db 110 KANODIATSKA 120
|||:|:|:|

RESULT 7
FLA_LEGMI STANDARD; PRT; 474 AA.
ID FLA_LEGMI
AC P53606;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Flagellin.
GN FLA.
OS Legionella micdadei.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Tallockia.
OX NCBI_TaxID=451;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96163446; PubMed=8562027;
RA Bangborg J., Hindersson P., Shand G., Holby N.;
RT "The Legionella micdadei flagellin: expression in Escherichia coli K
RT 12 and DNA sequence of the gene."
RL APMS 103:869-877(1995).
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X84699; CAA59172.1; -
DR InterPro: IPR001029; Flagellin_C.
DR InterPro: IPR001492; Flagellin_N.
DR Pfam: PF00700; Flagellin_C.1.
DR Pfam: PF00669; Flagellin_N.1.
DR PRINTS: PR00207; FLAGELLIN.
DR ProDom: PD000316; Flagellin_C.1.
KW Flagella.
SQ SEQUENCE 474 AA; 48113 MW; 5C9D05C8AEF9D07B CRC64;

Query Match 64.0%; Score 32; DB 1; Length 474;
Best Local Similarity 70.0%; Pred. No. 52;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 2 ASQDVTAVV 11
Db 211 ASQDATSAVA 220
|||:|:|:|

RESULT 8

```

HTR3_HALN1
ID HTR3_HALN1 STANDARD; PRT; 481 AA.
AC 09H000;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Halobacterial transducer protein III.
GN HTR9 OR VNG1395G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Ballea N.S., Thorsson V., Shroog A.J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weith R., Goo Y.A.,
RA Leitbauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Madocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Edhardt H., Lowe T.M., Liang P., Dennis P.P., Omer A.D.,
RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RT "Genome sequence of Halobacterium species NRC-1."
RL
CC -1- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC TRANSDUCTION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSOR DIMERIZATION DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE005058; ANG19717.1; -.
CC DR InterPro: IPR004089; Chemotaxis_transducer.
CC DR InterPro: IPR003660; HAMP.
CC DR InterPro: IPR004090; Me_Chemotaxis.
CC DR InterPro: IPR000014; PAS.
CC DR Pfam: PF00015; MCPsignal; 1.
CC DR PRINTS: PR00260; CHEMTRNSDUCR.
CC DR SMART: SM00304; HAMP; 1.
CC DR SMART: SM00283; MA; 1.
CC DR SMART: SM00091; PAS; 1.
CC DR PROSITE: PS50112; PAS; 1.
CC DR Transducer: Complete proteome.
CC KW Transducer: Complete proteome.
CC FT DOMAIN 10 81 PAS.
CC FT SEQUENCE 481 AA; 50773 MW; B33B55DBFE206D1 CRC64;
SQ

```

```

Query Match 64.0%; Score 32; DB 1; Length 481;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 3 SODVTTAVA 11
: :|:|:|:|:|
DB 447 ADDITTAVA 455

```

```

RESULT 9
HTR3_HALSA STANDARD; PRT; 481 AA.
ID HTR3_HALSA
AC 048316;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Halobacterial transducer protein III.
GN HTR9 OR HTP111.

```

```

OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-S9;
RX MEDLINE=96275896; PubMed=8674984;
RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA Oesterhelt D.;
RA "A family of halobacterial transducer proteins."
RL FEMS Microbiol. Lett. 139:161-168(1996).
CC -1- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC TRANSDUCTION.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSOR DIMERIZATION DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X95588; CA64840.1; -.
CC DR HSSP: P02942; 1007.
CC DR InterPro: IPR004089; Chemotaxis_transducer.
CC DR InterPro: IPR003660; HAMP.
CC DR InterPro: IPR000014; PAS.
CC DR Pfam: PF00015; MCPsignal; 1.
CC DR SMART: SM00304; HAMP; 1.
CC DR SMART: SM00283; MA; 1.
CC DR SMART: SM00091; PAS; 1.
CC DR PROSITE: PS50112; PAS; 1.
CC KW Transducer.
CC FT DOMAIN 10 81 PAS.
CC FT SEQUENCE 481 AA; 50728 MW; B283A873F7844A CRC64;
SQ

```

```

Query Match 64.0%; Score 32; DB 1; Length 481;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 3 SODVTTAVA 11
: :|:|:|:|:|
DB 447 ADDITTAVA 455

```

```

RESULT 10
KVLY_HUMAN STANDARD; PRT; 108 AA.
ID KVLY_HUMAN
AC P80362;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-1 region WAT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE AND X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=95086080; PubMed=7993911;
RA Huang D.-B., Chang C.-H., Ainsworth C., Bruenger A.T., Eulitz M.,
RA Solomon A., Stevens F.J., Schiffer M.;
RT "Comparison of crystal structures of two homologous proteins:
RT light-chain dimers."
RL Biochemistry 33:14848-14857(1994).
RN [2]
RP SEQUENCE OF 1-35.

```

RX MEDLINE-81267384; PubMed-6167731;
 RA Stevens F.J., Westholm F.A., Panagiotopoulos N., Schliffer M.,
 RA Popp R.A., Solomon A.,
 RT "Characterization and preliminary crystallographic data on the VL-
 RT Related fragment of the human ki Bence Jones protein Mat.,"
 RL J. Mol. Biol. 147:185-193(1981).
 CC -1- MISCELLANEOUS: THIS IS A BENGE-JONES PROTEIN.
 DR PDB: 1MTL; 01-NOV-94.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig; 1.
 DR SMART: SM00406; IGV; 1.
 KM Immunoglobulin V region; Bence-Jones protein; 3D-structure.
 FT DOMAIN 1 23 FRAMEWORK-1.
 FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.
 FT DOMAIN 35 49 FRAMEWORK-2.
 FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.
 FT DOMAIN 57 88 FRAMEWORK-3.
 FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.
 FT DOMAIN 98 107 FRAMEWORK-4.
 FT DISULFID 23 88 BY SIMILARITY.
 FT CONFLICT 30 31 TN -> SD (IN REF. 2).
 FT NON_TER 108 108
 SQ SEQUENCE 108 AA; 11737 MW; D9D941B3F0FAE697 CRC64;

Query Match 62.0%; Score 31; DB 1; Length 108;
 Best Local Similarity 60.0%; Pred. No. 18;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KASODVTAV 10
 Db 24 RASODITNYV 33

RESULT 11
 ID GYL2_HALN1 STANDARD; PRT; 279 AA.
 AC Q9HHU0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE GYPL protein 2.
 GN GYPL2 OR VNG6229G.
 OS Halobacterium sp. (strain NRC-1).
 OS plasmid pNRC200.
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=64091;
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20504483; PubMed-11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
 RA Shukla H.D., Lesky S.R., Baliga N.S., Thorsson V., Shroga J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Meddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohnschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1,"
 RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 CC -1- FUNCTION: MAY PLAY A STRUCTURAL OR REGULATORY ROLE IN GAS
 CC VESICLE SYNTHESIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE005156; AAC20886.1; ALT_INT.
 KM Gas vesicle; Plasmid; Complete proteome.
 SQ SEQUENCE 279 AA; 30376 MW; 543174A72EDF78B6 CRC64;
 Query Match 62.0%; Score 31; DB 1; Length 279;
 Best Local Similarity 66.7%; Pred. No. 49;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 ASODVTAV 10
 Db 206 ASODTTTSL 214

RESULT 12
 ID TCPA_CAEEL STANDARD; PRT; 549 AA.
 AC P41988; Q22228;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE T-complex protein 1, alpha subunit (TCP-1-alpha) (CCT-alpha).
 GN CCT-1 OR TCP-1 OR T05C12.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA MEDLINE-95278751; PubMed-7758963;
 RA Leroux M.R., Candido E.P.M.;
 RT "Molecular analysis of Caenorhabditis elegans tcp-1, a gene encoding
 RT a chaperonin protein,"
 RL Gene 156:241-246(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Burton J.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF PROTEINS UPON
 CC ATP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF
 CC ACTIN AND TUBULIN.
 CC -1- SUBUNIT: HEMERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 KDA THAT
 CC FORMS TWO STACKED RINGS, 12 TO 16 NM IN DIAMETER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U07941; AAB05072.1; -;
 DR EMBL: Z66500; CAA91308.1; -;
 DR HSSP: P48424; 1A6D.
 DR WormPep: T05C12.7; CE02319.
 DR InterPro: IPR002423; TCP1_cpn60.
 DR InterPro: IPR002194; TCP_1.
 DR Pfam: PF00118; cpn60_TCP1; 1.
 DR PRINTS: PR00304; TCOMPLEXTCP1.
 DR PROSITE: PS00750; TCP1_1; 1.
 DR PROSITE: PS00751; TCP1_2; 1.
 DR PROSITE: PS00995; TCP1_3; 1.
 KM Chaperone; ATP-binding; Multigene family.
 FT CONFLICT 469 469 F -> K (IN REF. 2).
 SQ SEQUENCE 549 AA; 58821 MW; D7C6D661FA391407 CRC64;

Query Match 62.0%; Score 31; DB 1; Length 549;

Best Local Similarity 77.88; Pred. No. 97;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 3 SODVTAVA 11
11:111111
Db 23 SONVATAVA 31

RESULT 13
ACON.CAEEL STANDARD; PRT; 788 AA.
ID ACON.CAEEL
AC P34455;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Probable aconitate hydratase, mitochondrial precursor (EC 4.2.1.3)
DE (Citrate hydro-lyase) (Aconitase).
GN P54H12.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
[1]
RN SQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sams M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Stalston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohlsman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- CATALYTIC ACTIVITY: Citrate - cis-aconitate + H(2)O.
CC -1- CORCTOR: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE (3FE-
CC 4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE CATALYTIC
CC SITE THAT INTERCONVERTS CITRATE, CIS-ACONITASE, AND ISOCITRATE (BY
CC SIMILARITY).
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial (by similarity).
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/TPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L25599; AAA8050.1; -
DR PIR: S44831; S44831.
DR HSP: P20004; 1AMJ.
DR WormPep: F54H12.1; CE00516.
DR InterPro: IPR001030; Aconitase.
DR InterPro: IPR000573; Aconitase_C.
DR Pfam: PF00330; aconitase_1.
DR Pfam: PF00694; aconitase_C; 1.
DR PRINTS: PR00415; ACONITASE.
DR PRODOM: PD000511; Aconitase; 1.
DR PROSITE: PS00450; ACONITASE_1; 1.
DR PROSITE: PS01244; ACONITASE_2; 1.
KM Hypothetical protein; lyase; tricarboxylic acid cycle; Iron-sulfur;
KW Mitochondrion; Transist peptide; 4fe-4s.
FT TRANSIT 1
FT MITOCHONDRION (POTENTIAL).

FT CHAIN ? 788 PROBABLE ACONITATE HYDRATASE.
FT METAL 393 393 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 456 456 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 459 459 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 788 AA; 85712 MW; 8861E6FC198B70D9 CRC64;

Query Match 62.0%; Score 31; DB 1; Length 788;
Best Local Similarity 66.7%; Pred. No. 1,4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 SODVTAVA 11
11:111111
Db 500 SPDITAMA 508

RESULT 14
EGT2_YEAST STANDARD; PRT; 1041 AA.
ID EGT2_YEAST
AC P42835;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE EGT2 protein precursor (Early G1 transcript 2).
GN EGT2 OR YNL327W OR N0320.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
[1]
RN SQUENCE FROM N.A.
RC STRAIN=S288C / FY1676;
RA MEDLINE=95373280; PubMed=7645347;
RA Mettani M., Nicoud J.-M., Levesque H., Gaillardin C.;
RT "Sequencing analysis of a 15.4 kb fragment of yeast chromosome XIV
RT identifies the RPD3, PAS8 and KRL1 loci, five new open reading
RT frames.";
RL yeast 11:567-572(1995).
RN [2]
RX SQUENCE FROM N.A., AND FUNCTION.
RP MEDLINE=96251274; PubMed=8668141;
RA Kovacech B., Nasmyth K., Schuster T.;
RT "EGT2 gene transcription is induced predominantly by Swi5 in early
RT G1.";
RL Mol. Cell. Biol. 16:3264-3274(1996).
CC -1- FUNCTION: SEEMS TO BE INVOLVED IN THE CORRECT TIMING OF CELL
CC SEPARATION AFTER CYTOKINESIS. AS SEPARATION OF MOTANT DAUGHTER
CC CELLS IS DELAYED, COULD EITHER BE AN ENZYME NECESSARY FOR GLUCANS-
CC DEGRADATION OF THE CELL WALL AT THE NECK REGION BETWEEN MOTHER AND
CC DAUGHTER CELLS OR A REGULATORY PROTEIN CONTROLLING THIS METABOLIC
CC STEP. EXCLUSIVELY EXPRESSED BETWEEN THE END OF MITOSIS AND EARLY
CC G1; INACTIVATED BEFORE CELLS PASS START.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z46259; CAAB6371.1; -
DR EMBL: Z71603; CAAG6259.1; -
DR SGD: S0005271; EGT2.
KW Glycoprotein; Signal; Cell cycle; Cell division; Mitosis.
FT SIGNAL 1 20
FT CHAIN 21 1041
FT DOMAIN 200 203 EGT2 PROTEIN.
FT DOMAIN 381 384 POLY-SER.
FT DOMAIN 388 395 POLY-SER.
FT DOMAIN 490 493 POLY-SER.
FT DOMAIN 586 589 POLY-THR.
FT REPEAT 457 492 1-1.

```
FT REPEAT 577 606 1-2.
FT REPEAT 613 647 1-3.
FT REPEAT 716 745 1-4.
FT REPEAT 773 802 1-5.
FT REPEAT 811 840 1-6.
FT REPEAT 849 886 1-7.
FT REPEAT 887 924 1-8.
FT REPEAT 925 962 1-9.
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 526 526 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 544 544 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 636 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1041 AA; 108494 MW; 01FECEFEAA8744CD CRC64;
```

```
Query Match 62.0%; Score 31; DB 1; Length 1041;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 KASQDVT 8
1:1:1:1:1
DB 111 KSSDIT 118
```

```
RESULT 15
YAU9_SCHPO
ID YAU9_SCHPO STANDARD; PRT; 1275 AA.
AC 010164;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 143.6 kDa protein C26A3.09C in chromosome I.
GN SPAC26A3.09C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: CONTAINS 1 PH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z69240; CA93232.1; -
CC HSSP; Q07960; IRGP.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000198; RHOGAP.
CC Pfam; PF00169; PH; 1.
```

```
DR Pfam; PF00620; RhogAP; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00324; RhogAP; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
KW Hypothetical protein
FT DOMAIN 719 836 PH.
SQ SEQUENCE 1275 AA; 143586 MW; 136DA1A26B4A1BBE CRC64;
```

```
Query Match 62.0%; Score 31; DB 1; Length 1275;
Best Local Similarity 63.6%; Pred. No. 2.3e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 KASQDVTAVA 11
1:1:1:1:1
DB 157 KSSKDVPTASA 167
```

```
Search completed: August 20, 2002, 11:41:20
Job time: 399 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:33:41 ; Search time 43.57 Seconds

(without alignments)
43.676 Million cell updates/sec

Title: US-09-824-286-4_COPY_24_34

Perfect score: 50
Sequence: 1 KASQDVTAAVA 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp_bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	38	76.0	812	2	Q9RBA5	Q9RBA5 arthropod
2	36	72.0	1449	2	O68542	O68542 streptococ
3	36	72.0	1449	2	O55264	O55264 streptococ
4	35	70.0	420	16	Q98NB5	Q98NB5 rhizobium 1
5	34	68.0	245	2	O9EUF7	O9EUF7 escherichia
6	34	68.0	264	10	Q9ATD1	O9ATD1 gossypium h
7	34	68.0	328	16	O927J6	O927J6 listeria in
8	34	68.0	389	2	O9I604	O9I604 salmonella
9	34	68.0	395	5	O17859	O17859 caenorhabdl
10	34	68.0	435	16	Q9A6D3	Q9A6D3 caulobacter
11	34	68.0	492	2	Q9AHT7	Q9AHT7 streptococ
12	34	68.0	501	2	O924J8	O924J8 streptococ
13	34	68.0	553	5	O9W458	O9W458 drosophila
14	34	68.0	584	2	O9KH13	O9KH13 mycoplasma
15	34	68.0	815	10	O80623	O80623 arabidopsis
16	34	68.0	1041	12	Q9E1F7	Q9E1F7 meleagrid h

ALIGNMENTS

17	34	68.0	1041	12	Q9DPQ2	Q9DPQ2 meleagrid h
18	34	68.0	2146	3	O59897	O59897 aspergillus
19	34	68.0	2146	3	O60026	O60026 aspergillus
20	33	66.0	226	5	P91481	P91481 caenorhabdl
21	33	66.0	400	4	O9NXH5	O9NXH5 homo sapien
22	33	66.0	449	4	O9H2K4	O9H2K4 homo sapien
23	33	66.0	449	4	O96F05	O96F05 homo sapien
24	33	66.0	485	12	O39258	O39258 equine herp
25	33	66.0	501	4	O960U6	O960U6 homo sapien
26	33	66.0	501	4	O96LX5	O96LX5 homo sapien
27	33	66.0	721	4	O9Y608	O9Y608 homo sapien
28	33	66.0	822	5	P91440	P91440 caenorhabdl
29	33	66.0	975	5	O9XZU1	O9XZU1 drosophila
30	33	66.0	975	5	O9VJH4	O9VJH4 drosophila
31	33	66.0	975	5	O9UB14	O9UB14 drosophila
32	33	66.0	2396	5	O77291	O77291 drosophila
33	32	64.0	117	10	O9LGC5	O9LGC5 oryza sativ
34	32	64.0	138	11	O9CPT6	O9CPT6 mus musculu
35	32	64.0	181	16	O98NF1	O98NF1 rhizobium 1
36	32	64.0	252	10	O9FJZ5	O9FJZ5 arabidopsis
37	32	64.0	263	16	O9HVT4	O9HVT4 pseudomonas
38	32	64.0	330	12	O69497	O69497 human herpe
39	32	64.0	348	12	O66597	O66597 epizootic h
40	32	64.0	355	5	O9VHT4	O9VHT4 drosophila
41	32	64.0	355	5	O9NFT2	O9NFT2 drosophila
42	32	64.0	371	10	O04342	O04342 leishmania
43	32	64.0	385	5	O9NLG2	O9NLG2 caenorhabdl
44	32	64.0	398	5	O27408	O27408 caenorhabdl
45	32	64.0	482	1	P71412	P71412 halobacteri

RESULT	1	PRELIMINARY;	PRT;	812 AA.
Q9RBA5				
AC	Q9RBA5			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	ENDO-INULINASE PRECURSOR (EC 3.2.1.7).			
GN	INU1.			
OS	Arthropod sp. S37.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Micrococccineae; Micrococccaceae; Arthropodacter.			
OX	NCBI_TaxID=85703;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-S37;			
RA	Kang S.I., Kim S.I.;			
RT	"Molecular cloning and sequence analysis of an endo-inulinase gene			
RT	from Arthropodacter sp. S37."			
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.			
CC	-I- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.			
DR	EMBL: AJ131562; CAB63119.1; "			
DR	Interpro: IPR001362; Glyco_hydro_32.			
DR	Pfam: PF00251; Glyco_hydro_32.			
KW	Glycosidase; Hydrolase; Signal.			
FT	CHAIN 1			
FT	SIGNAL 54			
FT	CHAIN 802			
SQ	SEQUENCE 812 AA; 87854 MW; 55A3E688A0C6B528 CRC64;			

Query Match 76.0%; Score 38; DB 2; Length 812;
Best Local Similarity 80.0%; Pred. No. 42;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KASQDVTAAV 10
Db 243 KASQDVTGTV 252

```
RESULT 2
ID 068542 PRELIMINARY; PRT; 1449 AA.
AC 068542;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE N (FRAGMENT).
GN GFEN.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V1477;
RA Jaffe R.I.;
RT "Streptococcus salivarius V1477 gtfN.";
RL Submitted (FEF-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF049609; AAC05156.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 8.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KM Transferase.
FT NON_TER
SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;
```

Query Match 72.0%; Score 36; DB 2; Length 1449;
Best Local Similarity 80.0%; Pred. NO. 1.9e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ASQDVTAVA 11
ID 69 AADPTTAVA 78

```
RESULT 3
ID 05264 PRELIMINARY; PRT; 1449 AA.
AC 05264;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE PRECURSOR.
GN GTF.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9512197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL: L35495; AAC41412.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 8.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KM Signal; Transferase.
FT SIGNAL
KW CHAIN
SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;
```

Query Match 72.0%; Score 36; DB 2; Length 1449;
Best Local Similarity 80.0%; Pred. NO. 1.9e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 ASQDVTAVA 11
ID 69 AADPTTAVA 78

```
RESULT 4
ID 098NB5 PRELIMINARY; PRT; 420 AA.
AC 098NB5;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE N-ACETYLMOBAMYOL-L-ALANINE AMIDASE.
GN MLR0213.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Matanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida T., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,
RA Mochizuki Y., Yamada M., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP002994; BAB47846.1; -.
DR InterPro: IPR002508; Amidase_3.
DR Pfam: PF01520; Amidase_3; 1.
KM Complete proteome.
SQ SEQUENCE 420 AA; 44793 MW; 9FAFA1B6A998B89 CRC64;
```

Query Match 70.0%; Score 35; DB 16; Length 420;
Best Local Similarity 63.6%; Pred. NO. 86;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KASQDVTAVA 11
ID 397 KAAQSTNAVA 407

```
RESULT 5
ID 09EUF7 PRELIMINARY; PRT; 245 AA.
AC 09EUF7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PILXS PROTEIN.
GN PILXS.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Nunez B.;
RL Thesis (1998), Universidad de Cantabria, Departamento de Biología.
DR EMBL: AJ006342; CAC20142.1; -.
KM Plasmid.
SQ SEQUENCE 245 AA; 27705 MW; FB253DC89C368155 CRC64;
```

Query Match 68.0%; Score 34; DB 2; Length 245;
Best Local Similarity 54.5%; Pred. NO. 78;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 KASQDVTAVA 11
 Db 174 KATODATNAIS 184

RESULT 6

Q9ATD1 PRELIMINARY; PRT; 264 AA.
 AC Q9ATD1: 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 GN GHMYB9.
 OS Gossypium hirsutum (Upland cotton).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Malvales; Malvaceae; Gossypium.
 OX NCBI_TaxID=3635;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ACALA MAXX; TISSUE=DAY OF ANTHESIS OVULE;
 RA Matz E.C., Burr B.;
 RT "Cotton seed fibers are trichomes."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF336286; AAK19619.1; -
 DR HSSP: P01103; 1POM.
 DR InterPro: IPR001005; MYD_DNA_bind.
 DR Pfam: PF00249; MYD_DNA-binding; 2.
 DR SMART; SM00395; SANT; 2.
 DR PROSITE; PS00037; MYB_1; 1.
 DR PROSITE; PS00334; MYB_2; 1.
 DR PROSITE; PS50090; MYB_3; 2.
 SQ SEQUENCE 264 AA; 29629 MW; E09BCB4E7662E95 CRC64;

Query Match 68.0%; Score 34; DB 10; Length 264;
 Best Local Similarity 87.5%; Pred. No. 84;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVT 8
 Db 131 EASQDVT 138

RESULT 7
 Q927J6 PRELIMINARY; PRT; 328 AA.
 AC Q927J6: 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 GN LIN2792. PROTEIN.
 OS Listeria innocua.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Listeria.
 OX NCBI_TaxID=1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 11262 / SEROVAR 6A;
 RA Glaeser P., Frangoul L., Buchrieser C., Rusnock C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Chabrit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussutget O.,
 RA Eutlian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurupkati G.,
 RA Madeno E., Maitournan A., Mata Vicente J., Ng E., Negjari H.,
 RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;

RT "Comparative genomics of Listeria species."
 RL Science 294:849-852(2001).
 DR EMBL: AL596173; CAC98018.1; -
 DR ListList: LIN02792; -
 KW Complete proteome.
 SQ SEQUENCE 328 AA; 36670 MW; F2E1A0F2C6D98622 CRC64;

Query Match 68.0%; Score 34; DB 16; Length 328;
 Best Local Similarity 87.5%; Pred. No. 11e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 SODVTAV 10
 Db 153 SEDVTAV 160

RESULT 8
 Q9L6O4 PRELIMINARY; PRT; 389 AA.
 AC Q9L6O4: 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE 89% IDENTIFY WITH E. COLI PUTATIVE UROPOPHRYNGIN GEN III METHYLASE
 (HEMX) (SP:p09127).
 GN HEMX.
 OS Salmonella typhimurium LT2.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=99287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SGSC1412;
 RA Wastu;
 RT "The Salmonella typhimurium Genome Sequencing Project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SGSC1412;
 RA Waterston R.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF233324; AAF33455.1; -
 KW Methyltransferase.
 SQ SEQUENCE 389 AA; 42285 MW; 25B73FA79157770E CRC64;

Query Match 68.0%; Score 34; DB 2; Length 389;
 Best Local Similarity 70.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 ASQDVTAVA 11
 Db 151 SDQDVTAAA 160

RESULT 9
 O17859 PRELIMINARY; PRT; 395 AA.
 AC O17859: 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 GN F32H2.7. PROTEIN.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kershaw J.K.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "genome sequence of the nematode *C. elegans*: A platform for
 RL investigating biology";
 RL Science 282:2012-2018(1998).
 DR EMBL: Z81523; CAB04247.1; -.
 SQ SEQUENCE 395 AA; 43071 MW; 814EB82A7BE1160D CRC64;

Query Match 68.0%; Score 34; DB 5; Length 395;
 Best Local Similarity 70.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KASODVTTAV 10
 Db 365 RSSODVLTAV 374

RESULT 10

09A6D3 ID 09A6D3 PRELIMINARY; PRT; 435 AA.
 AC 09A6D3;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.
 GN CC2161.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=69394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CH15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nielsen W.C., Feldidylum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eiken J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Ptoczka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Ueberlack T., Tran K., Wolf A., Vamathavan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL: AEO05888; AAK24132.1; -.
 DR MEROPS: S11.0PW: -.
 DR TIGR: CC2161;
 DR InterPro: IPR001967; DadacBptasel.
 DR Pfam: PF00768; Peptidase_S11; 1.
 DR PRINTS: PR00725; DADACBPTASE1.
 KW Carboxypeptidase; Complete proteome.
 SQ SEQUENCE 435 AA; 46511 MW; 8A0DBD1971495B3E CRC64;

Query Match 68.0%; Score 34; DB 16; Length 435;
 Best Local Similarity 63.6%; Pred. No. 1.4e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KASODVTTAVA 11
 Db 132 KSANDVATAVA 142

RESULT 11

09AHT7 ID 09AHT7 PRELIMINARY; PRT; 492 AA.
 AC 09AHT7;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 1,4-BETA-N-ACETYLMURAMIDASE (FRAGMENT).

GN LYTC.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M4;
 RX MEDLINE=21116976; PubMed=11179332;
 RA Witzmann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
 RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
 RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
 RA Langemann S., Johnson S., Koenig S.;
 RT "Use of a whole Genome Approach To Identify Vaccine Molecules
 Affording Protection against *Streptococcus pneumoniae* Infection.";
 RL Infect. Immun. 69:1593-1598(2001).
 DR EMBL: AF291697; AAK19157.1; -.
 FT NON_TER 1 1
 SQ SEQUENCE 492 AA; 57554 MW; DF6DC1F1DA33B043 CRC64;

Query Match 68.0%; Score 34; DB 2; Length 492;
 Best Local Similarity 77.8%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KASODVTTA 9
 Db 31 KTSODVTTA 39

RESULT 12

09Z4J8 ID 09Z4J8 PRELIMINARY; PRT; 501 AA.
 AC 09Z4J8;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE 1,4-BETA-N-ACETYLMURAMIDASE PRECURSOR (EC 3.2.1.17).
 GN LYTC.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M31;
 RX MEDLINE=99340545; PubMed=10411730;
 RA Garcia P., Gonzalez M.P., Garcia E., Garcia J.L., Lopez R.;
 RT "The molecular characterization of the first autolytic lysozyme of
 RT Streptococcus pneumoniae reveals evolutionary mobile domains.";
 RL Mol. Microbiol. 33:128-138(1999).
 DR EMBL: AJ009639; CAA08765.1; -.
 KW Signal; Hydrolase; Glycosidase.
 FT SIGNAL 1 33
 FT CHAIN 34 501
 SQ SEQUENCE 501 AA; 58682 MW; A6774B6DBF2EB704 CRC64;

Query Match 68.0%; Score 34; DB 2; Length 501;
 Best Local Similarity 77.8%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KASODVTTA 9
 Db 40 KTSODVTTA 48

RESULT 13

09W458 ID 09W458 PRELIMINARY; PRT; 553 AA.
 AC 09W458;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)
 DE CG12236 PROTEIN.
 GN CG12236.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aspayani A., An H.-J., Andrews-Planckoch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Flossler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,
 RA Jalali M., Kalush F., Kapran G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostreli A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AEO03435; AAF46100.1; -;
 DR FlyBase: FBgn0029822; CG12236.
 DR InterPro: IPR000210; BTB_POZ.
 DR InterPro: IPR000822; Znf-C2H2.
 DR Pfam: PF00651; BTB; 1.
 DR Pfam: PF00096; zf-C2H2; 2.
 DR SMART: SM00225; BTB; 1.
 DR SMART: SM00355; ZNF_C2H2; 2.
 DR PROSITE: PS50097; BTB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
 KW DNA-binding; Metal-binding; Zinc-finger.
 SO SEQUENCE 553 AA; 60801 MW; EB3019D282819092 CRC64;

Query Match 68.0%; Score 34; DB 5; Length 553;
 Best Local Similarity 63.6%; Pred. No. 1.9e+02;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 KASODVTTAVA 11
 DB 346 QTSODATTA 356

RESULT 14
 ID 09KH13 PRELIMINARY; PRT; 584 AA.
 AC 09KH13:
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)
 DE ADHESIN PMGAL.4 (FRAGMENT).
 GN PMGAL.4.
 OS Mycoplasma gallisepticum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 NCBI_Taxid=2096;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HS;
 RA Shen Q.C., Bi D.R., Weng C.J.;
 RT "Sequence analysis of the pmga multigene family of Mycoplasma
 RT gallisepticum strain HS.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF275312; AAF91415.1; -;
 DR InterPro: IPR001986; EPSP_synthase.
 DR PROSITE: PS00104; EPSP_SYNTHASE_1; UNKNOWN.1.
 FT NON_TER 584
 SO SEQUENCE 584 AA; 62453 MW; 5C467BA55FB27A72 CRC64;

Query Match 68.0%; Score 34; DB 2; Length 584;
 Best Local Similarity 54.5%; Pred. No. 1.9e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 KASODVTTAVA 11
 DB 224 KANEDIMTAVS 234

RESULT 15
 ID 080623 PRELIMINARY; PRT; 815 AA.
 AC 080623:
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, last annotation update)
 DE PUTATIVE SER/THR PROTEIN KINASE (HYPOTHETICAL 91.3 KDA PROTEIN).
 GN F12I6.2 OR ATG39360.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_Taxid=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
 RA Somerville C.R., Venter J.C.;
 RT "Arabidopsis thaliana chromosome II BAC F12I6 genomic sequence.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Theologis A., Ecker J., Davis R.W.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL: AC004218; AAC27827.1; -;
 DR EMBL: AF325084; AAK17152.1; -;
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR004040; STY_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR SMART: SM00221; STYKC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;
KW Transferase.
SQ SEQUENCE 815 AA; 91321 MW; D3F4E07B73D8342D CRC64;

Query Match 68.0%; Score 34; DB 10; Length 815;
Best Local Similarity 87.5%; Pred. NO. 2.7e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ASQDVTTA 9
|||:|
Db 110 ASQDVTTA 117

Search completed: August 20, 2002, 11:40:53
Job time: 432 sec

FE	FT	FT	FT	FT	FT	FT	FT	PN	XX	XX	XX	XX	XX
FE	FT	FT	FT	FT	FT	FT	FT	PN	XX	XX	XX	XX	XX
09-MAY-1997;	Region	Region	Region	Region	Region	Region	Region	MO9343416-A1.	20-NOV-1997.	24..34	50..56	89..97	97WO-US07870.
										"/note="	"/note="	"/note="	
										"CDR1"	"CDR2"	"CDR3"	

PR 10-MAY-1996; 96US-0017466.
XX (BIOJ) BIOGEN INC.
PA
XX Benjamin CD, Burkly LC, Hession C, Whitty A;
PI
XX WPI: 1998-008885/01.
DR N-PSDB: AAT97440.
XX
XX Blocking agents of the gamma common chain of cytokine receptors -
PT particularly monoclonal antibodies, used to induce T cell energy for
PT treatment of immunological diseases
XX
XX Claim 22: Page 81; 111pp; English.
PS
XX This polypeptide comprises the light chain variable region (VL) of
CC monoclonal antibody (Mab) C9.B8, which is produced by a hybridoma
CC deposited as ATCC 12107, and which is specific for the gamma
CC constant (gc) chain (see AAW31646) of human cytokine receptors. The
CC invention provides compositions and methods for inhibiting cytokine
CC signalling using gc chain blocking agents for the treatment of
CC immunological diseases such as myasthenia gravis, rheumatoid
CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,
CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,
CC allergy, asthma, parasitic infection, graft vs. host disease or
CC psoriasis. Preferred gc blocking agents include Mab CP.B8, its Fab
CC fragment and an antibody having a light chain variable region
CC CDR selected from those of CP.B8 VL or a heavy chain variable
CC region CDR selected from those of CP.B8 VH (see AAW31646).
XX
SQ Sequence 106 AA;

Query Match 100.0%; Score 42; DB 19; Length 106;
Best Local Similarity 100.0%; Pred. NO. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7
| | | | | | |
Db 50 wastrht 56

RESULT 2
AAR25723
ID AAR25723 standard; Protein: 107 AA.
XX
AC AAR25723;
XX

DT 13-JAN-1993 (first entry)
XX

DE Humanised VL region of the mouse Fd 138-80 antibody.
XX

XX Murine: immunoglobulin; CDR; HSV; non immunogenic; herpes simplex
KW virus; HSV; light chain; variable region; framework; human; Eu.
XX

OS Mus musculus.
XX

XX Key Location/Qualifiers
FH Region 24..38
FT /note- "CDR"

FT Region 50..56
FT /note- "CDR"

FT Region 89..97
FT /note- "CDR"

FT Misc-difference 36
FT /note- "mutated residue"

FT Misc-difference 48
FT /note- "mutated residue"

FT Misc-difference 63
FT /note- "mutated residue"

FT Misc-difference 87
FT /note- "mutated residue"

FT
XX

PN W09211018-A.
XX
XX 09-JUL-1992.
PD
XX
XX 19-DEC-1991; 91WO-US09711.
PF
XX 19-DEC-1990; 90US-0634278.
PR
XX (PROT-) PROTEIN DESIGN LABS INC.
PA
XX Co MS, Coellnigh KL, Landolf NF, Queen CL, Schneider WP;
PI
XX WPI: 1992-249842/30.
DR
XX
XX New immunoglobulin(s) having murine CDRs in human framework
PT regions - have lower antigenicity; useful for treating e.g. HSV,
PT CMV, T-cell disorders, myeloid disorders and auto-immune
PT conditions
XX
XX Claim 19: Fig 14D; 141pp; English.
PS

XX The sequence shows the humanised mature light chain variable region
CC of the mouse Fd 138-80 antibody. Murine CDRs were used in a
CC human Eu framework to produce a pure humanised immunoglobulin (Ig)
CC which is capable of binding to a herpes simplex virus-specific
CC epitope. The Ig is non immunogenic, due to the human framework,
CC and has a strong affinity for its predetermined antigen. They can
CC be produced in large quantities via recombinant DNA and monoclonal
CC antibody technology. The humanised Igs may be used alone or in
CC combination with chemotherapeutic agents such as non-steroidal
CC anti-inflammatory drugs or immunosuppressants.
CC See also AAR25721-32.
XX

SQ Sequence 107 AA;

Query Match 100.0%; Score 42; DB 13; Length 107;
Best Local Similarity 100.0%; Pred. NO. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7
| | | | | | |
Db 50 wastrht 56

RESULT 3
AAR55121
ID AAR55121 standard; Protein: 107 AA.
XX
AC AAR55121;
XX

DT 27-JAN-1995 (first entry)
XX

DE Mouse anti-HIV mu39.1 light chain variable region.
XX

XX Immunoglobulin; heavy chain; anti-HIV antibody; neutralisation;
KW human immunodeficiency virus; variable region; VH chain; murine.
XX

OS Mus musculus.
XX

XX Key Location/Qualifiers
FH Region 1..23
FT /label- FR1

FT Region 24..34
FT /label- CDR1

FT Region 35..49
FT /label- FR2

FT Region 50..56
FT /label- CDR2

FT Region 57..88
FT /label- FR3

FT Region 89..97
FT /label- CDR3

```

FT Region 98..107.
FT /label= FR4
XX
XX JP06125783-A.
XX
XX 10-MAY-1994.
XX
XX 28-DEC-1991; 91JP-0359808.
XX
XX 28-DEC-1991; 91JP-0359808.
XX
XX (KAGA-) 2H KAGAKU KESSEI-RYOHO KENKYUSHO.
XX
XX WPI; 1994-187942/23.
XX
XX N-PSDB; AAO65552.
XX
XX Mouse-human chimeric anti-HIV antibody heavy and light chains -
XX PT and recombinant antibody consisting of the H- and L-chains,
XX PT useful in AIDS therapy
XX
XX Example 3; Fig 2; 22pp; Japanese.
XX
XX Murine monoclonal antibodies mu 39.1 and mu 5.5 both neutralise HIV.
XX CC The heavy and light chain variable regions from these antibodies
XX CC were sequenced (AAR55120-R55123). The murine anti-HIV CDRs were
XX CC introduced into human framework regions to construct chimeric
XX CC antibodies (AAR55124-R55127).
XX
XX Sequence 107 AA:

```

```

Query Match 100.0%; Score 42; DB 15; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 WASTRHT 7
DB 50 wastrht 56

```

```

RESULT 4
AAR55125 ID AAR55125 standard; Protein; 107 AA.
XX
XX AAR55125;
XX
XX 30-JAN-1995 (first entry)
XX
XX Mouse-human chimeric anti-HIV mu39.1-derived light chain V region.
XX
XX Immunoglobulin; heavy chain; anti-HIV antibody; neutralisation;
XX KW human immunodeficiency virus; variable region; VL chain; murine;
XX KW chimeric; humanised.
XX
XX Chimeric Mus musculus.
XX OS Chimeric Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT 1..23 /label= FR1
XX FT /note= "human"
XX FT 24..34 /label= CDR1
XX FT /note= "murine"
XX FT 35..49 /label= FR2
XX FT /note= "human, except for murine residue Ser49"
XX FT 50..56 /label= CDR2
XX FT /note= "murine"
XX FT 57..88 /label= FR3
XX FT /note= "human"

```

```

FT Region 89..97
FT /label= CDR3
FT /note= "murine"
XX
XX Region 98..107
XX
XX /label= FR4
XX /note= "human"
XX
XX JP06125783-A.
XX
XX 10-MAY-1994.
XX
XX 28-DEC-1991; 91JP-0359808.
XX
XX 28-DEC-1991; 91JP-0359808.
XX
XX (KAGA-) 2H KAGAKU KESSEI-RYOHO KENKYUSHO.
XX
XX WPI; 1994-187942/23.
XX
XX N-PSDB; AAO65556.
XX
XX Mouse-human chimeric anti-HIV antibody heavy and light chains -
XX PT and recombinant antibody consisting of the H- and L-chains,
XX PT useful in AIDS therapy
XX
XX Claim 5; Fig 10; 22pp; Japanese.
XX
XX Murine monoclonal antibodies mu 39.1 and mu 5.5 both neutralise HIV.
XX CC The heavy and light chain variable regions from these antibodies
XX CC were sequenced (AAR55120-R55123). The murine anti-HIV CDRs were
XX CC introduced into human framework regions to construct chimeric
XX CC antibodies (AAR55124-R55127).
XX
XX Sequence 107 AA:

```

```

Query Match 100.0%; Score 42; DB 15; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 WASTRHT 7
DB 50 wastrht 56

```

```

RESULT 5
AAR60300 ID AAR60300 standard; Protein; 107 AA.
XX
XX AAR60300;
XX
XX 09-MAR-1995 (first entry)
XX
XX Anti HIV antibody light chain variable region.
XX
XX Antibody; heavy chain; light chain; human immunodeficiency virus;
XX KW HIV; acquired immune deficiency syndrome; AIDS; treatment;
XX KW prophylaxis; Mus musculus; Homo sapiens.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers
XX FT 1..23 /label= Framework region 1.
XX FT 24..34 /label= CDR1.
XX FT 35..49 /label= Framework region 2.
XX FT 50..56 /label= CDR2.
XX FT 57..88 /label= Framework region 3.
XX FT 89..97 /label= CDR3.

```

FT	Region	98..107
FT	/label=	Framework region 4.
XX		
PN	MO9415969-A.	
XX		
PD	21-JUL-1994.	
XX		
PE	14-JAN-1993;	93WO-JP00039.
XX		
PR	14-JAN-1993;	93AU-0032671.
XX		
PR	14-JAN-1993;	93WO-JP00039.
XX		
PA	(KAGA)	CHEMO SERO THERAPEUTIC RES INST.
XX		
PI	Eda Y, Kimachi K, Maeda H, Osatomi K, Shiosaki K;	
PI	Tokiyoshi S;	
XX		
DR	WPI: 1994-249145/30.	
DR	N-PSDB; AAQ70370.	
XX		
PT	Recombinant chimeric anti HIV antibody - useful for the treatment	
PT	and prevention of HIV	
XX		
PS	Claim 7; Figure 2; 51pp; Japanese.	
XX		
CC	The recombinant antibody light chain has neutralising activity	
CC	against HIV. Chimeric antibodies comprising both mouse and human	
CC	sequences are useful in the treatment/prevention of AIDS caused by	
CC	HIV.	
XX		
SO	Sequence	107 AA;
	Query Match	100.0%; Score 42; DB 15; Length 107;
	Best Local Similarity	100.0%; Pred. No. 0.72;
	Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 WASTRHT 7	
DB	50 wastrht 56	
RESULT	6	
AA060304		
ID	AA060304 standard; Protein: 107 AA.	
XX		
AC	AA060304;	
XX		
DT	13-MAR-1995 (first entry)	
XX		
DE	Chimeric anti HIV antibody light chain variable region.	
XX		
KW	Antibody; heavy chain; light chain; human immunodeficiency virus;	
KW	HIV; acquired immune deficiency syndrome; AIDS; treatment;	
KW	prophylaxis; Mus musculus; Homo sapiens.	
XX		
OS	Chimeric Homo sapiens.	
OS	Chimeric Mus musculus.	
XX		
XX		
Key	Location/Qualifiers	
FT	Region	1..23
FT	/label=	Framework region 1.
FT	Region	24..34
FT	/label=	CDR1.
FT	/note=	"Mouse derived amino acid sequence."
FT	Region	35..49
FT	/label=	Framework region 2.
FT	Region	49..56
FT	/note=	"Mouse derived amino acid sequence."
FT	Region	50..56
FT	/label=	CDR2.
FT	Region	57..88
FT	/label=	Framework region 3.

```

FT      Region          89..97      /label- CDR3.
FT      /note- "Mouse derived amino acid sequence."
FT      Region          98..107     /label- Framework region 4.
XX      MO9415969-A.
XX      PD      21-JUL-1994.
XX      PE      14-JAN-1993;      93WO-JP00039.
XX      PR      14-JAN-1993;      93AU-0032671.
XX      PR      14-JAN-1993;      93WO-JP00039.
XX      (KAGA ) CHEMO SERO THERAPEUTIC RES INST.
XX      XX      Eda Y, Kimachi K, Maeda H, Osatomi K, Shiosaki K;
PI      Tokiyoshi S;
XX      WPI: 1994-249145/30.
DR      N-PSDB; AAQ70374.
XX      PT      Recombinant chimeric anti HIV antibody - useful for the treatment
PT      and prevention of HIV
XX      PS      Claim 6; Figure 10; 51pp; Japanese.
XX      CC      The recombinant antibody light chain has neutralizing activity
CC      against HIV. Chimeric antibodies comprising both mouse and human
CC      sequences are useful in the treatment/prevention of AIDS caused by
CC      HIV.
XX      SO      Sequence      107 AA;

Query Match          100.0%; Score 42; DB 15; Length 107;
Best Local Similarity 100.0%; Pred. NO. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 WASTRHT 7
      111111
Db      50 wastrht 56

RESULT      7
AAW47087
ID      AAW47087 standard; Protein; 107 AA.
XX      AAW47087;
XX      AC
XX      DT      26-JUN-1998 (first entry)
DE      Mouse J591 monoclonal antibody light chain variable region VK17.
XX      DE
KW      Mouse; monoclonal antibody; J591; prostate specific membrane antigen;
KW      cancer; vascular endothelial cell; metastatic adenocarcinoma.
XX      OS      Mus sp.
XX      PN      WO9803873-A1.
XX      PD      29-JAN-1998.
XX      PE      17-JUL-1997;      97WO-US12035.
XX      PR      09-APR-1997;      97US-0838682.
XX      PR      18-JUL-1996;      96US-0022125.
XX      PA      (CORR ) CORNELL RES FOUND INC.
XX      PI      Bander NH;
XX

```

DR WPI: 1998-120937/11.
DR N-PSDB: AAV13953.

PT Destroying cancer cells with agent that binds to prostate specific
PT membrane antigen - on vascular endothelial cells near the cancer, or
PT on normal, hypertrophic or cancerous prostatic cells, also used for
PT diagnosis

PS Example 12: Page 61; 94pp; English.

XX The present sequence represents the mouse J591 monoclonal antibody light
CC chain variable region from clone VK17 from an example of the present
CC invention. The present invention describes the elimination of cancer
CC cells by treating vascular endothelial cells (VEC) close to the cancer
CC with an agent (A) able to bind to the extracellular domain (ECD) of
CC prostate specific membrane antigen (PSMA). (A) both binds to the VEC and
CC destroys the cancer cells. Also described are: (1) the detection of
CC cancer tissue by detecting binding of labelled (A) to VEC close to, or
CC within, a cancer tissue; (2) eliminating or detecting normal, benignly
CC hyperplastic or cancerous prostate epithelial cells using optionally
CC labelled (A); (3) hybridomas that produce a monoclonal antibody (Mab)
CC that binds to PSMA. The method is used to treat renal, urothelial,
CC colon, lung, rectal or breast cancers and metastatic adenocarcinoma of
CC the liver. The diagnostic method is particularly used to detect
CC recurrence of prostatic disease or to monitor the effect of treatments
CC for prostate cancer (presence of PSMA in the serum indicates that
CC prostate cells are being lysed). (A) binds to an epitope of PSMA
CC expressed on live cells (contrast antibody 7B11 which only binds after
CC cell lysis), allowing targeting of live, unfixed cells and thus
CC providing more efficient treatment and diagnosis. Both cancer cells
CC themselves and the VEC on which they depend are killed. All VEC close to
CC cancer cells express PSMA, whatever the type of cancer, but normal VEC
CC do not.

SQ Sequence 107 AA;

Query Match 100.0%; Score 42; DB 19; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7
| | | | |
Db 50 wastrht 56

RESULT 8

ID AAY90375 standard; Protein; 107 AA.

XX AAY90375;

DT 15-JAN-2001 (first entry)

XX J591 monoclonal antibody light chain protein sequence fragment.

DE J591 monoclonal antibody: extracellular domain; diagnosis: therapy;

KW J591 monoclonal antibody: extracellular domain; diagnosis: therapy;

KM prostate specific membrane antigen; prostate cancer; light chain.

XX Homo sapiens.

OS US6107090-A.

XX 22-AUG-2000.

XX 09-APR-1997; 97US-0838682.

XX 06-MAY-1996; 96US-0016976.

XX 18-JUL-1996; 96US-0022125.

XX (CORR) CORNELL RES FOUND INC.

XX Bander NH;

XX WPI: 2000-571325/53.
DR N-PSDB: AAA37835.

PT Antibody specific for extracellular prostate-specific membrane antigen,
PT useful for diagnosis and treatment of prostate cancer

PS Example 12: Column 24; 33pp; English.

XX This sequence is a fragment of the light chain of the monoclonal antibody
CC J591. The invention relates to an isolated antibody or its antigen
CC binding portion (I) which binds to an extracellular domain of prostate
CC specific membrane antigen and which does not require cell lysis to bind
CC to the extracellular domain. The antibody or its antigen binding portion
CC is selected for its ability to bind to live cells. (I) is useful for
CC diagnosis of diseases associated with the presence of normal, benign
CC hyperplastic, and cancerous epithelial cells or portions. Also it can be
CC used for identifying the recurrence of such diseases, particularly when
CC the disease is localised in a particular biological material of the
CC patient for e.g. recurrence of prostatic disease. They can also be used
CC alone or bound to a substance effective to kill cancerous prostate
CC epithelial cells as a therapy for prostate cancer. Binding and
CC internalising of the antibody with the prostate specific membrane
CC antigen, permits the therapeutic use of intracellularly acting cytotoxic
CC agents. (I) targets only prostate epithelial cells and other tissue are
CC spared which provides safer treatment particularly for elderly patients.
CC The antibodies bind to living prostate cells and treatments using these
CC antibodies are more effective than those which target lysed prostate
CC cells.

SQ Sequence 107 AA;

Query Match 100.0%; Score 42; DB 21; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7
| | | | |
Db 50 wastrht 56

RESULT 9

ID AAB69665 standard; Protein; 107 AA.

XX AAB69665;

DT 30-APR-2001 (first entry)

XX Murine Fd138-80 antibody light chain SEQ ID NO: 50.

KW Humanised immunoglobulin; mouse; human; antibody; heavy chain; diabetes;

KM light chain; graft versus host disease; transplant; autoimmune disease;

KW multiple sclerosis; rheumatoid arthritis; systemic lupus erythematosus;

KM myasthenia gravis; herpes infection; myeloid leukaemia; CMV infection.

XX Mus sp.

OS US6180370-B1.

XX 30-JAN-2001.

XX 07-JUN-1995; 95US-0484537.

XX 28-DEC-1988; 88US-0290975.

XX 13-FEB-1989; 89US-0310252.

XX 28-SEP-1990; 90US-0590274.

XX 19-DEC-1990; 90US-0634278.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Queen CL, Selick HE;

XX WPI: 2001-190856/19.
XX
XX Producing humanized immunoglobulin, involves producing a cell
PT containing DNA segments encoding humanized heavy and light chain
PT variable regions, and expressing the DNA segments in the cell -
XX
XX Example 6; Fig 3; 145pp; English.
XX
XX The present invention describes a method of producing humanised
CC immunoglobulins involving expressing in a cell a nucleic acid encoding a
CC humanised version of an immunoglobulin. This is obtained by comparing a
CC donor and human immunoglobulin and producing a combined antibody which
CC contains part of each. These are useful in the treatment of
CC graft-versus-host disease, transplant rejection, autoimmune diseases such
CC as diabetes, rheumatoid arthritis, myasthenia gravis, multiple sclerosis
CC and systemic lupus erythematosus, herpes infections, CMV virus infections
CC and myeloid leukaemia. The present sequence is an antibody used to
CC demonstrate the method of the invention.
XX
XX Sequence 107 AA;
SQ

Query Match 100.0%; Score 42; DB 22; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WASTRHT 7
| | | | |
Db 50 wastrht 56

RESULT 10
AAB69666
ID AAB69666 standard; Protein; 107 AA.
XX

AC AAB69666;
XX

DT 30-APR-2001 (first entry)
XX

DE Humanised Fd138-80 antibody light chain SEQ ID NO: 51.
XX

KW Humanised immunoglobulin; mouse; human; antibody; heavy chain; diabetes;
KW light chain; graft versus host disease; transplant; autoimmune disease;
KW multiple sclerosis; rheumatoid arthritis; systemic lupus erythematosus;
KW myasthenia gravis; herpes infection; myeloid leukaemia; CMV infection.
XX

OS Mus sp.
OS Homo sapiens.
XX

PN US6180370-B1.
XX

PD 30-JAN-2001.
XX

PE 07-JUN-1995; 95US-0484537.
XX

PR 28-DEC-1988; 88US-0290975.
XX

PR 13-FEB-1989; 89US-0310252.
XX

PR 28-SEP-1990; 90US-0590274.
XX

PR 19-DEC-1990; 90US-0634278.
XX

PA (PROT-) PROTEIN DESIGN LABS INC.
XX
XX Queen CL, Selick HE;
XX
XX WPI: 2001-190856/19.
XX

PT Producing humanized immunoglobulin, involves producing a cell
PT containing DNA segments encoding humanized heavy and light chain
PT variable regions, and expressing the DNA segments in the cell -
XX
XX Example 6; Fig 3; 145pp; English.
XX

CC The present invention describes a method of producing humanised
CC immunoglobulins involving expressing in a cell a nucleic acid encoding a
CC humanised version of an immunoglobulin. This is obtained by comparing a
CC donor and human immunoglobulin and producing a combined antibody which
CC contains part of each. These are useful in the treatment of
CC graft-versus-host disease, transplant rejection, autoimmune diseases such
CC as diabetes, rheumatoid arthritis, myasthenia gravis, multiple sclerosis
CC and systemic lupus erythematosus, herpes infections, CMV virus infections
CC and myeloid leukaemia. The present sequence is an antibody used to
CC demonstrate the method of the invention.
XX
XX Sequence 107 AA;
SQ

Query Match 100.0%; Score 42; DB 22; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 WASTRHT 7
| | | | |
Db 50 wastrht 56

RESULT 11
AAB36226
ID AAB36226 standard; Protein; 107 AA.
XX

AC AAB36226;
XX

DT 15-FEB-2001 (first entry)
XX

DE Monoclonal antibody J591 kappa light chain V region #2.
XX

KW Mouse; antibody; heavy chain; prostate cancer; biological agent.
XX

OS Mus sp.
XX

PN US6136311-A.
XX

PD 24-OCT-2000.
XX

PE 17-JUL-1997; 97US-0895914.
XX

PR 06-MAY-1996; 96US-0016976.
XX

PR 18-JUL-1996; 96US-0022125.
XX

PR 09-APR-1997; 97US-0838682.
XX

PA (CORR) CORNELL RES FOUND INC.
XX

PI Bander NH;
XX

DR WPI: 2001-040234/05.
XX

DR N-PSDB; AAC66546.
XX

XX
XX Example 12; Column 26; 35pp; English.
XX

CC The present invention describes a method of killing cancer cells,
CC particularly prostate cancer cells, by directing a biological agent to
CC the cells which then binds to a prostate specific membrane antigen and
CC causes the molecule to be internalised. The internalisation of the agent,
CC which may be bound to a drug or which may act to kill the cell alone,
CC then leads to the death of the cell. The present sequence forms
CC part of an antibody which may be used as the biological agent of the
CC invention. In addition to prostate cancer, the method can be used with
CC renal, urothelial, colon, renal, lung and breast cancer cells, and
CC cancerous cells of metastatic adenocarcinoma to the liver.
XX

SQ Sequence 107 AA;
XX

Query Match 100.0%; Score 42; DB 22; Length 107;
 Best Local Similarity 100.0%; Pred. No. 0.72; Mismatches 7; Conservative 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7
 |||||
 Db 50 wastrht 56

RESULT 12
 AAY21817
 ID AAY21817 standard; Protein; 108 AA.
 XX
 AC AAY21817;
 XX
 DT 10-SEP-1999 (first entry)
 XX
 DE Anti-STX1 light chain variable region.
 XX
 KM Humanised; monoclonal antibody; Mab: Shiga toxin; immunoglobulin;
 KM Ig constant region; enterohemorrhagic Escherichia coli; EHEC; uremia;
 KM edema; bloody diarrhoea; hemorrhagic colitis; hemolytic uremic syndrome;
 KM thrombocytopenia; EHEC-mediated disease; anti-STX1.
 XX
 OS Shigella dysenteriae.
 XX
 FH Location/Qualifiers
 FT Region 24..34
 FT /note= "complementarity determining region (CDR) 1;
 FT specifically claimed for in claim 10"
 FT Region 50..56
 FT /note= "complementarity determining region (CDR) 2;
 FT specifically claimed for in claim 10"
 FT Region 89..97
 FT /note= "complementarity determining region (CDR) 3;
 FT specifically claimed for in claim 10"
 XX
 PN MO9932645-A1.
 XX
 PD 01-JUL-1999.
 XX
 PF 22-DEC-1998; 98MO-0527267.
 XX
 PR 18-DEC-1998; 98US-0215163.
 PR 23-DEC-1997; 97US-0068635.
 XX
 PA (MELT/) MELTON-CELSA A.
 PA (OBRI/) O'BRIEN A D.
 PA (SCHM/) SCHMITT C K.
 PA (STIN/) STINSON J L.
 PA (WONG/) WONG H.
 XX
 PI Melton-Celsa A, O'Brien AD, Schmitt CK, Stinson JL;
 PI Wong H;
 XX
 DR WPI: 1999-418935/35.
 DR N-PSDB: AAX82029.
 XX
 XX Humanized monoclonal antibodies against Shiga toxins, useful for
 PT protection against enterohemorrhagic Escherichia coli or other Shiga
 PT toxin producing bacteria
 XX
 PS Claim 5; Fig 3; 75pp; English.
 XX
 CC The invention relates to humanised monoclonal antibodies (MAB) against
 CC Shiga toxins. The humanised MAB that binds to Shiga toxin comprises a
 CC constant and a variable region, where: (a) the constant region contains
 CC at least part of a human immunoglobulin (Ig) constant region; and (b) the
 CC variable region contains at least part of a non-human Ig variable region.
 CC Host cells transformed with vectors encoding a humanised MAB against
 CC Shiga toxin type 2 is useful for treating a patient with an infection

CC caused by enterohemorrhagic Escherichia coli (EHEC) or other Shiga toxin
 CC producing bacteria. The humanised MAB can also be used to reduce illness
 CC caused by EHEC or other Shiga toxin producing bacteria. EHEC are
 CC associated with food-borne outbreaks of bloody diarrhoea (hemorrhagic
 CC colitis) and the hemolytic uremic syndrome. In particular, the humanised
 CC MABs ameliorate edema, thrombocytopenia and uremia associated with EHEC-
 CC mediated disease. The present sequence represents an anti-STX1 light
 CC chain variable region.
 XX
 SQ Sequence 108 AA;

Query Match 100.0%; Score 42; DB 20; Length 108;
 Best Local Similarity 100.0%; Pred. No. 0.73; Mismatches 7; Conservative 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7
 |||||
 Db 50 wastrht 56

RESULT 13
 AAP40031
 ID AAP40031 standard; Protein; 223 AA.
 XX
 AC AAP40031;
 XX
 DT 01-DEC-1991 (first entry)
 XX
 DE Kappa anti-carcinoembryonic antigen chain.
 XX
 KM Immunoglobulin; carcinoembryonic antigen.
 XX
 PN EP125023-A.
 XX
 PD 14-NOV-1984.
 XX
 PF 06-APR-1984; 84EP-0302368.
 XX
 PR 08-APR-1983; 83US-0483457.
 XX
 PA (CITY) CITY OF HOPE.
 PA (GETH) GENENTECH INC.
 XX
 PI Cabilly S, Holmes WE, Wetzel RB, Heyneker HL, Riggs AD;
 XX
 DR WPI: 1984-283749/46.
 DR N-PSDB: AANA0023.
 XX
 PT Immunoglobulin(s) produced by recombinant host cells - useful as
 PT antibodies analogous to forms from mammals.
 XX
 PS Disclosure; Fig. 3; 79pp; English.
 XX
 CC The protein is encoded by the mRNA coding sequence of the cDNA fragment
 CC contained within recombinant vector pK17G4, using the vector this
 CC immunoglobulin is produced readily in pure monoclonal form. Genetic
 CC manipulations can be used to produce chimeras of variants drawing their
 CC homology from species differing from each other. Protein manipulation
 CC is also possible.
 XX
 SQ Sequence 223 AA;

Query Match 100.0%; Score 42; DB 5; Length 223;
 Best Local Similarity 100.0%; Pred. No. 1.5; Mismatches 7; Conservative 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7
 |||||
 Db 59 wastrht 65

RESULT 14
ID AAP80155 standard; protein; 247 AA.
XX
AC AAP80155;
XX
DT 13-OCT-1990 (first entry)
XX
DE FRA and CDRs from anti-CEA monoclonal antibody.
XX
KW Biosynthetic Antibody Binding site (BABS); variable domain;
anti-CEA monoclonal antibody; ss.
XX
OS Synthetic.
XX
PN WO8809344-A.
XX
PD 01-DEC-1988.
XX
PE 19-MAY-1988; 88WO-US01737.
XX
PR 21-MAY-1987; 87US-0052800.
XX
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Huston JS, Oppermann H;
XX
DR WPI: 1988-353928/49.
DR N-PSDB; AAN80181.
XX
PT Recombinant multifunctional protein - having an Ab binding site and a
sequence for biological activity, ion sequestering or binding to a solid
support.
XX
PS Disclosure: ; 15pp; English.
XX
CC Multi-functional biosynthetic protein comprising FRS and
CC CDRs with specificity for a tumour antigen CEA.
CC See also AAN80171-N80192.
XX
SQ Sequence 247 AA;
QY 1 WASTRHT 7
DB 190 wastrht 196

Query Match 100.0%; Score 42; DB 9; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 15
ID AAP80156 standard; protein; 247 AA.
XX
AC AAP80156;
XX
DT 14-NOV-1990 (first entry)
XX
DE Biosynthetic antibody binding site.
XX
KW Biosynthetic antibody binding site; framework region; assay; imaging;
multifunctional protein.
XX
PN WO8809344-A.
XX
PD 01-DEC-1988.
XX
PE 19-MAY-1988; 88WO-US01737.
XX
PR 21-MAY-1987; 87US-0052800.
XX

PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Huston JS, Oppermann H;
XX
DR WPI: 1988-353928/49.
DR N-PSDB; AAN80182.
XX
PT Recombinant multifunctional protein - having antibody binding site and a
sequence for biological activity, ion sequestering or binding to a
solid support.
XX
PS Disclosure: ; 15pp; English.
XX
CC The biosynthetic antibody binding site forms part of a recombinant
CC multifunctional protein which also comprises an effector protein, an AA
CC sequence capable of sequestering an ion, or a sequence capable of
CC binding to a solid support. The BABS comprises framework regions in both
CC VH and LH and CDR1 and CDR3 in VH, and CDR1, CDR2 and CDR3 in VL from an
CC anti-CEA monoclonal antibody. CDR2 in VH is a CDR2 consensus sequence
CC found in most immunoglobulin VH regions. The effector protein is an
CC enzyme, toxin, receptor, binding site, growth factor, cytokine
CC or antimetabolite. The sequence capable of sequestering an ion is
CC calmodulin or metallothionein. The sequence able to bind to a solid
CC support is streptavidin or a protein A fragment. The protein may be used
CC for, eg specific binding assays, affinity purification, biocatalysts, drug
CC targeting, imaging and immunological treatment of oncogenic and
CC infectious diseases. The protein offers fewer cleavage sites to
CC circulating proteolytic enzymes and have improved stability.
CC They reach target organs rapidly and are cleared quickly from the body.
XX
SQ Sequence 247 AA;
QY 1 WASTRHT 7
DB 190 wastrht 196

Query Match 100.0%; Score 42; DB 9; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: August 20, 2002, 11:34:39
Job time: 633 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:35:17 ; Search time 25.74 Seconds
(without alignments)
26.132 Million cell updates/sec

Title: US-09-824-286-4_COPY_50_56

Perfect score: 42
Sequence: 1 WASTRHT 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	117	2 S42466	Ig kappa chain V r
2	42	100.0	131	2 P10207	anti-idiotypic ant
3	42	100.0	152	2 S30751	Ig kappa chain pre
4	34	81.0	212	2 T27834	hypothetical prote
5	34	81.0	246	2 C38113	yebC protein - Esc
6	34	81.0	246	2 F90950	hypothetical prote
7	34	81.0	246	2 B85799	hypothetical prote
8	34	81.0	246	2 AG0743	conserved hypotnet
9	34	81.0	605	2 T49780	related to beta tr
10	34	81.0	3010	1 GNMWTW	genome polypotein
11	33	78.6	206	2 C83590	probable transcript
12	33	78.6	212	2 C90786	probable tet opero
13	33	78.6	212	2 C64843	probable tet opero
14	33	78.6	212	2 A85646	probable tet opero
15	33	78.6	275	2 D97525	hypothetical prote
16	33	78.6	347	2 D83554	hypothetical prote
17	33	78.6	380	2 S62525	hypothetical prote
18	33	78.6	1139	2 S44800	FlOe9.8 protein -
19	32	76.2	142	2 C82829	conserved hypotnet
20	32	76.2	268	2 E86962	probable oxidoredu
21	32	76.2	268	2 E70551	probable oxidoredu
22	32	76.2	434	2 AG1693	hypothetical prote
23	32	76.2	572	2 DEECPC	pyruvate dehydrog
24	32	76.2	572	2 E90748	pyruvate oxidase I
25	32	76.2	572	2 A85599	pyruvate oxidase P
26	32	76.2	725	2 D75294	hypothetical prote
27	32	76.2	738	2 E75354	ethylene-response
28	32	76.2	745	2 E75354	DNA helicase II -
29	32	76.2	889	2 A41259	potassium transpor

30	32	76.2	1330	1 VCBES5	major capsid prote
31	32	76.2	1385	2 S61236	major capsid prote
32	31	73.8	99	2 C46518	Ig L1 chain V regi
33	31	73.8	115	2 B46518	Ig L1 chain V regi
34	31	73.8	171	2 AC2539	hypothetical prote
35	31	73.8	196	2 G84513	hypothetical prote
36	31	73.8	223	2 A91083	hypothetical prote
37	31	73.8	223	2 E65059	hypothetical prote
38	31	73.8	223	2 C85928	hypothetical prote
39	31	73.8	223	2 AH0409	conserved hypotnet
40	31	73.8	223	2 AB0859	conserved hypotnet
41	31	73.8	238	2 AB6562	hypothetical prote
42	31	73.8	238	2 C72061	conserved hypotnet
43	31	73.8	238	2 D71513	hypothetical prote
44	31	73.8	238	2 AB1671	conserved hypotnet
45	31	73.8	247	2 AG0250	conserved hypotnet

ALIGNMENTS

RESULT 1
S42466
Ig kappa chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
C:Accession: S42466
R:Shlyanov, P.A.; Beaspalov, I.A.; Terletskaya, H.N.; Deyev, S.M.
Submitted to the EMBL Data Library, March 1994
A:Reference number: S42466
A:Accession: S42466
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-117 <SHT>
A:Cross-references: EMBL:X78108, NID:9460824, PIDN:CMA54998.1, PID:9460825
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: heterotetramer; Immunoglobulin
F:26-100/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 42; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 100.0%; Score 42; DB 2; Length 131;

RESULT 2
P10207
anti-idiotypic antibody E225, kappa chain V region precursor - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 21-Jan-2000
C:Accession: P10207
R:Souchon, H.; Doyen, N.; Riottot, M.M.; Rougeon, F.; Poljak, R.J.
MOL. Immunol. 27, 429-433, 1990
A:Title: Nucleotide sequence of the VH, VL regions of an anti-idiotypic antibody rea
A:Reference number: P10207; MUID:90309764
A:Accession: P10207
A:Molecule type: mRNA
A:Residues: 1-131 <SOU>
A>Note: This sequence corresponds to subgroup V mouse immunoglobulin light chain
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
F:1-20/Domain: signal sequence #status predicted <SIG>
F:36-110/Domain: immunoglobulin homology <IMM>
F:44-54/Region: complementarity-determining 1
F:70-76/Region: complementarity-determining 2
F:109-117/Region: complementarity-determining 3
F:116-127/Region: J region

Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7
|||||
Db 70 WASTRHT 76

RESULT 3
S30751
Ig kappa chain precursor V region - mouse
C:Species: Mus musculus (house mouse)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Jan-2000
C:Accession: S30751
R:Grant, F.J.; Levin, S.D.; Gilbert, T.; Kindsvogel, W.
Nucleic Acids Res. 15, 5496, 1987
A:Title: Improved RNA sequencing method to determine immunoglobulin mRNA sequence.
A:Reference number: S30751; MUID:87260030
A:Accession: S30751
A:Molecule type: mRNA
A:Residues: 1-152 <GRA>
A:CROSS-references: EMBL:X05877; NID:952195; PIDN:CAA29301.1; PID:952196
C:Keywords: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:45-119/Domain: immunoglobulin homology <IMM>
F:136-132/Domain: C region (C-kappa) (fragment) #status predicted <CRE>

Query Match
Best Local Similarity 100.0%; Score 42; DB 2; Length 152;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7
|||||
Db 79 WASTRHT 85

RESULT 4
T27834
hypothetical protein ZK384.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T27834
R:Ainscough, R
Submitted to the EMBL Data Library, November 1996
A:Reference number: Z20427
A:Accession: T27834
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-212 <WLI>
A:CROSS-references: EMBL:Z82092; PIDN:CA805011.1; GSPDB:GN00023; CESP:ZK384.2
C:Genetics:
A:Gene: CESP:ZK384.2
A:Map position: 5
A:Introns: 60/3; 106/3; 180/2
C:Superfamily: yellowjacket venom allergen antigen 5

Query Match
Best Local Similarity 81.0%; Score 34; DB 2; Length 212;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 6
|||||
Db 145 WASTRHT 150

RESULT 5
C38113
yebc protein - Escherichia coli
C:Species: Escherichia coli
C>Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 20-Jun-2000

C:Accession: C38113; S19014; H64948
R:Takehagi, M.; Iwasaki, H.; Nakata, A.; Shinagawa, H.
J. Bacteriol. 173, 5747-5753, 1991
A:Title: Molecular analysis of the Escherichia coli ruvC gene, which encodes a Hollid
A:Reference number: A38113; MUID:91358366
A:Accession: C38113
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-246 <TA>
A:CROSS-references: GB:D10165; GB:D90392; NID:9216649; PIDN:BAA01031.1; PID:9216652
R:Sharples, G.J.; Lloyd, R.G.
J. Bacteriol. 173, 7711-7715, 1991
A:Title: Resolution of Holliday junctions in Escherichia coli: identification of the
A:Reference number: S19013; MUID:92041688
A:Accession: S19014
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <SHA>
A:CROSS-references: EMBL:X59551; NID:942172; PIDN:CAA42126.1; PID:942173
A:Experimental source: strain K-12
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H64948
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-246 <BLAT>
A:CROSS-references: GB:AE000280; GB:U00096; NID:91788163; PIDN:AAC74934.1; PID:917881
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yebc
C:Superfamily: hypothetical protein MG332

Query Match
Best Local Similarity 81.0%; Score 34; DB 2; Length 246;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 6
|||||
Db 7 WASTRHT 12

RESULT 6
F90950
hypothetical protein ECs2574 [Imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: F90950
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90950
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <HAY>
A:CROSS-references: GB:BA000007; PIDN:BA835997.1; PID:913362042; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs2574
C:Superfamily: hypothetical protein MG332

Query Match
Best Local Similarity 81.0%; Score 34; DB 2; Length 246;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 6
|||||

Db 7 WANTRH 12

RESULT 7
B85799

hypothetical protein yebc [imported] - Escherichia coli (strain O157:H7, substrain EDL93)

C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85799

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85799

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <STO>
A:Cross-references: GB:AE005174; MID:g12515920; PIDN:AGS56854.1; GSPDB:GN00145; UWGP:Z29
A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:
A:Gene: yebc
C:Superfamily: hypothetical protein MG332

Query Match 81.0%; Score 34; DB 2; Length 246;
Best Local Similarity 83.3%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6
11:111
Db 7 WANTRH 12

RESULT 8
AG0743

conserved hypothetical protein STY2107 [imported] - Salmonella enterica subsp. enterica

C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG0743

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
Th, T.; Connelton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moulie, S.; O'Gaora, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AG0743

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05650.1; PID:g16503146; GSPDB:GN00176
C:Genetics:
A:Gene: STY2107
C:Superfamily: hypothetical protein MG332

Query Match 81.0%; Score 34; DB 2; Length 246;
Best Local Similarity 83.3%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6
11:111
Db 7 WANTRH 12

RESULT 9
T49780

related to beta transducin-like protein [imported] - Neurospora crassa

N:Alternate names: protein B9J10.170
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C:Accession: T49780
R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Farthmann, B.; Holland, R.; Nykati
submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022
A:Accession: T49780

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-605 <SCH>
A:Cross-references: EMBL:AL356324; GSPDB:GN00116; NCSP:B9J10.170
A:Experimental source: BAC clone B9J10, strain OR74A

C:Genetics:
A:Gene: NCSP:B9J10.170
A:Map position: 6
A:Introns: 183/2

Query Match 81.0%; Score 34; DB 2; Length 605;
Best Local Similarity 71.4%; Pred. No. 33;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRHT 7
11:111
Db 201 WASTRHT 207

RESULT 10
GNMWYM

genome polyprotein - hepatitis C virus (strain Taiwan)

N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstr
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244

R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992

A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping
A:Reference number: A40244; MUID:92230206
A:Accession: A40244

A:Molecule type: genomic RNA
A:Residues: 1-3010 <CHE>
A:Cross-references: GB:M84754

C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: Arg; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc
E:1-115/Product: capsid protein C #status predicted <GPC>
E:116-101/Product: envelope protein M #status predicted <PPM>
E:192-389/Product: major envelope protein E #status predicted <MEP>
E:730-729/Product: nonstructural protein NS1 #status predicted <NS1>
E:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
E:1007-1615/Product: hepatitisin #status predicted <NS3>
E:1230-1237/Region: nucleotide-binding motif A (P-loop)
E:1312-1317/Region: nucleotide-binding motif B
E:1316-1319/Region: DEXH motif
E:1616-1862/Product: nonstructural protein NS4a #status predicted <NA4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NA5>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041

Query Match 81.0%; Score 34; DB 1; Length 3010;
Best Local Similarity 71.4%; Pred. No. 1,8e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRHT 7
11:111
Db 2816 WASTRHT 2822

RESULT 11
C83590

probable transcription regulator PA0436 [imported] - Pseudomonas aeruginosa (strain
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

R:Accession: C83590
C:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, J.; Lory, S.; Olson, M.V.
N:ature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337
A:Accession: C83590
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <STO>
A:Cross-references: GB:AE004481; GB:AE004091; NID:g9946293; PIDN:AG03825.1; GSPDB:GN00101
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0436

Query Match	78.6%	Score 33	DB 2	Length 206	
Best Local Similarity	83.3%	Pred. No. 17			
Matches	5	Conservative	1	Mismatches	0
				Indels	0
				Gaps	0
Qy	1	WASTRH	6		
		:			
Db	160	WASTRH	165		

```

RESULT      12
C90786
Probable test operon regulator [imported] - Escherichia coli (strain O157:H7, substrain F
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence.revision 18-Jul-2001 #text.change 18-Jul-2001
C:Accession: C90786
R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genoc
A:Reference number: A99629, MUID:21156231, PMID:11258796
A:Accession: C90786
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA034682.1; PID:g13360719; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics:
:Gene: ECs1259

```

```

Query March 78.6%; Score 33; DB 2; Length 212;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6
||||:|
Db 167 WASTOH 172

RESULT 13
C64843
Probable transcription regulator ydcC - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: C64843
R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:9742617
A:Accession: C64843
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-212 <BLAST>
A:Cross-references: GB:AE00203; GB:U00096; MID:g1787248; PIDN:AACT4098.1; PID:g1787249;
A:Experimental source: strain K-12, substrain MK1655

```

C:Genetics:
A:gene: ycdc
C:Keywords: DNA binding; transcription regulation
F:39-58/Region: helix-turn-helix motif

	Query Match	78.6%	Score 33;	DB 2;	Length 212;
	Best Local Similarity	83.3%	Pred. No. 18;		
	Matches	5;	Conservative	1;	Mismatches
				0;	Indels
					Gaps
Oy	1 WASTRH	6			
		:			
Db	167 WASTOH	172			

RESULT 14
 A:56546
 probable tet operon regulator ycdC [imported] - Escherichia coli (strain O157:H7, sub
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 14-Sep-2001
 C:Accession: A85646
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apodaca
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: A85646
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-212 <STO>
 A:Cross-references: GB:A005174; NID:q12514373; PIDN:AAG55629.1; GSPDB:GN00145; UMGCP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: ycdC

```

Query Match      78.6%;   Score 33;   DB 2;   Length 212;
Best Local Similarity 83.3%;   Pred. No. 18;
Matches 5;   Conservative 1;   Mismatches 0;   Indels 0;   Gaps 0;

QY 1 WASTRH 6
    | | | | |
Db 167 WASTQH 172

```

RESULT 15
D97525
hypothetical protein AGR_C_2525 [imported] - Agrobacterium tumefaciens (strain C58, C
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: D97525
R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A.: Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11745194
A:Accession: D97525
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-275 <RUR>
A:Cross-references: GB:AE007869; PIDN:AAK87157.1; PID:g15156429; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2525
A:Map position: circular chromosome

Query Match	78.6%	Score 33	DB 2	Length 275
Best Local Similarity	71.4%	Pred. No. 24		
Matches	5	Conservative	1	Mismatches 1
				Indels 0
				Gaps 0
OY	1	WASTRHT	7	
db	212	WASARHS	218	

Search completed: August 20, 2002, 11:35:18
Job time: 397 sec

2002 AUG 20 12:58:55

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:41:20 ; Search time 15.95 Seconds

(without alignments)
16.993 Million cell updates/sec

Title: US-09-824-286-4_COPY_50_56

Perfect score: 42

Sequence: 1 WASTRHT 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	81.0	246	1	YEBC_ECOLI
2	34	81.0	3010	1	POLG_HCVTW
3	33	78.6	212	1	YCDB_ECOLI
4	33	78.6	380	1	YC9B_SCHPO
5	33	78.6	1207	1	YI08_CAEEL
6	32	76.2	572	1	POXB_ECOLI
7	32	76.2	889	1	TRK2_YEAST
8	32	76.2	1330	1	VCAP_PRVIS
9	31	73.8	238	1	Y457_CHLTR
10	31	73.8	238	1	Y573_CHLPR
11	31	73.8	238	1	Y742_CHLMU
12	31	73.8	251	1	Y003_MYCLE
13	31	73.8	251	1	Y003_MYCLE
14	31	73.8	445	1	SCRC_RABIT
15	31	73.8	449	1	SCRC_RABIT
16	31	73.8	461	1	PPBA_BACSU
17	31	73.8	461	1	PPBA_BACSU
18	31	73.8	461	1	PPBA_BACSU
19	31	73.8	502	1	SYE_MYCLE
20	31	73.8	994	1	SYV_XYLFK
21	31	73.8	1228	1	C1BA_BACTK
22	30	71.4	114	1	KV4A_HUMAN
23	30	71.4	121	1	KV40_HUMAN
24	30	71.4	133	1	KV4B_HUMAN
25	30	71.4	134	1	KV4C_HUMAN
26	30	71.4	378	1	MUTY_HAEN
27	30	71.4	552	1	YGBB_BACSU
28	30	71.4	566	1	KADS_SCHPO
29	30	71.4	675	1	PRTS_BOVIN
30	30	71.4	1229	1	C1BB_BACTU
31	30	71.4	1231	1	C1BD_BACTZ
32	30	71.4	1233	1	C1BC_BACTM
33	29	69.0	75	1	YCCJ_ECOLI

34	29	69.0	166	1	PEPI_PYRPU
35	29	69.0	181	1	ORN_VIBCH
36	29	69.0	232	1	FKBL_ARATH
37	29	69.0	259	1	DEOC_VIBCH
38	29	69.0	263	1	CTR2_CANFA
39	29	69.0	342	1	GALR_SALTU
40	29	69.0	343	1	GALR_ECOLI
41	29	69.0	387	1	DADR_MOUSE
42	29	69.0	387	1	DADR_MOUSE
43	29	69.0	392	1	CGR2_RATIA
44	29	69.0	420	1	OC3A_DROME
45	29	69.0	488	1	YGM9_YEAST

ALIGNMENTS

RESULT ID	1	YEBC_ECOLI	STANDARD	PRT	246 AA.
AC	P24237				
DT	01-MAR-1992 (Rel. 21, Created)				
DT	01-MAR-1992 (Rel. 21, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Protein yebc.				
GN	YEBC OR B1864.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID:562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91358366; PubMed=1885548;				
RA	Takahagi M., Iwasaki H., Nakata A., Shingawa H.;				
RT	"Molecular analysis of the Escherichia coli ruvc gene, which encodes a Holliday junction-specific endonuclease.";				
RT	J. Bacteriol. 173:7711-7715(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91358366; PubMed=1885548;				
RA	Takahagi M., Iwasaki H., Nakata A., Shingawa H.;				
RT	"Molecular analysis of the Escherichia coli ruvc gene, which encodes a Holliday junction-specific endonuclease.";				
RT	J. Bacteriol. 173:5747-5753(1991).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=K12 / MG1655.				
RX	MEDLINE=97426617; PubMed=9278503;				
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,				
RT	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,				
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,				
RT	Mau B., Siao Y.;				
RT	"The complete genome sequence of Escherichia coli K-12.";				
RT	Science 277:1453-1474(1997).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN=K12;				
RX	MEDLINE=97251358; PubMed=9097040;				
RA	Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,				
RA	Isono K., Kasai H., Kimura S., Kitagawa M., Kitagawa M.,				
RA	Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,				
RA	Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,				
RA	Saito N., Sampei G., Seki Y., Sivasubram S., Tagam H.,				
RA	Takeda J., Takemoto K., Wada C., Yamamoto Y., Horikuchi T.;				
RT	"A 460-kb DNA sequence of the Escherichia coli K-12 genome				
RT	corresponding to the 40.1-50.0 min region on the linkage map.";				
RT	DNA Res. 3:379-392(1996).				
RN	[5]				
RP	IDENTIFICATION BY MASS SPECTROMETRY.				
RX	MEDLINE=99420866; PubMed=10493123;				
RA	Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;				
RT	"Enrichment of low abundance proteins of Escherichia coli by				
RT	hydroxyapatite chromatography.";				

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D53CDE215 CRC64;

Query Match
 Best Local Similarity 71.4%; Pred. No. 78;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WASTRHT 7
 |||:
 Db 2816 WATARHT 2822

RESULT 3
 YC9B_ECOLI STANDARD; PRT; 212 AA.
 ID YC9B_ECOLI
 AC P75899;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical transcriptional regulator ycdC.
 GN YC9B OR B1013.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 mln region on the linkage map";
 RL DNA Res. 3:137-155(1996)
 CC -1- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL
 REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000203; AAC74098.1; -
 DR EMBL: D90738; BAA35790.1; -
 DR Ecogene; EG12301; ycdC.
 DR InterPro; IPR001647; HTH_Tetr.
 DR Pfam; PF00440; tetr. 1.
 DR PROSITE; PS01081; HTH_TETR_FAMILY; FALSE_NEG.
 KM Hypothetical protein; Transcription regulation; DNA-binding;
 KW Complete proteome.
 FT DNA_BIND 39 58 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 212 AA; 23687 MW; E2459B85DFAC277A CRC64;

Query Match
 Best Local Similarity 78.6%; Score 33; DB 1; Length 212;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WASTRH 6
 |||:
 Db 167 WASTOH 172

RESULT 4
 YC9B_SCHPO STANDARD; PRT; 380 AA.
 ID YC9B_SCHPO
 AC O09885;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Hypothetical 43.0 kDa protein C584.11C in chromosome III.
 GN SPCC584.11C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL032824; CAB37424.1; -
 DR Hypothetical protein.
 KW SEQUENCE 380 AA; A65777B672310E68 CRC64;
 SQ SEQUENCE 380 AA; A65777B672310E68 CRC64;

Query Match
 Best Local Similarity 78.6%; Score 33; DB 1; Length 380;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WASTRHT 7
 |||:
 Db 178 WASKMHT 184

RESULT 5
 YL08_CAEEL STANDARD; PRT; 1207 AA.
 ID YL08_CAEEL
 AC P34402;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Hypothetical 138.6 kDa protein F10E9.8 in chromosome III.
 GN F10E9.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copley T., Cooper J., Coulson A.,
 RA Craxton J., Dear S., Du Z., Durbin R., Favello A., Fraser A.,

RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopla A., Saunders D., Shownkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L10986; AAA28012.2; -
DR WormPep: F10E9.8; CE00557.
KW Hypothetical protein.
SQ SEQUENCE 1207 AA; 138631 MW; EB13FBC52AC0A3FC CRC64;

Query Match 78.6%; Score 33; DB 1; Length 1207;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WASTRH 6
| | | | |
DB 164 WGSTRH 169

RESULT 6
POXB_ECOLI STANDARD; PRT; 572 AA.
ID POXB_ECOLI
AC P07003; Q47513; Q47514; Q47515; Q47516; Q47517; Q47518; Q47519;
AC Q47520;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pyruvate dehydrogenase [cytochrome] (EC 1.2.2.2) (Pyruvate oxidase)
DE (POX) (Pyruvate dehydrogenase [ubiquinone]).
GN POXB OR B0871.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-K12;
RX MEDLINE-86286555; PubMed-3016647;
RA Grabau C., Cronan J.E. Jr.;
RT "Nucleotide sequence and deduced amino acid sequence of Escherichia
RT coli pyruvate oxidase, a lipid-activated flavoprotein.";
RL Nucleic Acids Res. 14:5449-5460(1986).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97061202; PubMed-8905232;
DR Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,

RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horichi T.;
RT "A 718-bp DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
[4]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RC STRAIN-K12;
RX MEDLINE-89308683; PubMed-2663858;
RA Grabau C., Chang Y.Y., Cronan J.E. Jr.;
RT "Lipid binding by Escherichia coli pyruvate oxidase is disrupted by
RT small alterations of the carboxyl-terminal region.";
RL J. Biol. Chem. 264:12510-12519(1989).
[5]
RP SEQUENCE OF 550-572 FROM N.A.
RX MEDLINE-86033917; PubMed-3902830;
RA Rechy M.A., Grabau C., Cronan J.E. Jr., Hager L.P.;
RT "Characterization of the alpha-peptide released upon protease
RT activation of pyruvate oxidase.";
RL J. Biol. Chem. 260:14287-14291(1985).
[6]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE-94293772; PubMed-8022274;
RA Chang Y.Y., Wang A.Y., Cronan J.E. Jr.;
RT "Expression of Escherichia coli pyruvate oxidase (POXB) depends on
RT the sigma factor encoded by the rpoS(katF) gene.";
RL Mol. Microbiol. 11:1019-1028(1994).
CC -1- CATALYTIC ACTIVITY: Pyruvate + ferriocytochrome b1 + H(2)O = CO(2)
CC + acetate + ferriocytochrome b1.
CC -1- COFACTOR: THIAMINE PYRROPHOSPHATE, FAD AND MAGNESIUM.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- PTM: ACTIVATED BY LIMITED PROTEOLYTIC DIGESTION. THIS CLEAVAGE
CC PRODUCES A PEPTIDE (ALPHA-PEPTIDE) AND MIMICS THE ACTIVATION OF
CC ENZYME BY PHOSPHOLIPIDS. THE PROTEOLYTIC CLEAVAGE ALSO RESULTS IN
CC THE LOSS OF THE HIGH AFFINITY LIPID-BINDING SITE OF THE ENZYME.
CC -1- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL: X04105; CAA27725.1; -
EMBL: AE000186; AAC73958.1; -
DR EMBL: D90724; BAA35585.1; -
DR EMBL: S73268; AAB31180.1; -
DR EMBL: M28208; AAB59101.1; -
DR EMBL: L47688; AAB59102.1; -
DR EMBL: L47689; AAB59103.1; -
DR EMBL: L47690; AAB59104.1; -
DR EMBL: L47691; AAB59105.1; -
DR EMBL: L47692; AAB59106.1; -
DR EMBL: L47693; AAB59107.1; -
DR EMBL: L47694; AAB59108.1; -
DR EMBL: L47695; AAB59109.1; -
PIR: A23648; DEECPC.
DR HSSP: P37063; IPOM.
DR SWISS-2DPAGE: P07003; COLI.
DR ECO2DBASE: G058.0; 6TH EDITION.
DR Ecocode: EG10754; POXB.
DR InterPro: IPR000399; TPP_enzyme.
DR Pfam: PF00205; TPP_enzymes; 1.
DR Pfam: PF02775; TPP_enzymes-C; 1.
DR Pfam: PF02776; TPP_enzymes-N; 1.
DR PROSITE: PS00187; TPP_ENZYMES; 1.
KW Oxidoreductase; Flavoprotein; FAD; Thiamine pyrophosphate; Magnesium;

KW Membrane; Lipid-binding: Complete proteome.
 FT PEPTIDE 550 572 ALPHA-PEPTIDE.
 FT ACT_SITE 50 50 BY SIMILARITY.
 FT MUTAGEN 533 533 A->T: IN POXB11.
 FT MUTAGEN 553 553 A->V: IN POXB14.
 FT MUTAGEN 560 560 D->P: IN POXB15; NORMAL ACTIVITY.
 FT MUTAGEN 564 564 E->P: IN POXB16; LOSS OF ACTIVITY.
 FT MUTAGEN 572 572 R->G: IN POXB10; REDUCED ACTIVITY. MAY
 FT MUTAGEN 549 572 INTERACT LESS WITH MEMBRANES.
 FT MUTAGEN 564 572 MISSING: IN POXB6.
 FT MUTAGEN 570 572 MISSING: IN POXB7.
 FT CONFLICT 364 365 MISSING: IN POXB8.
 FT CONFLICT 414 416 OO -> HE (IN REF. 4).
 SO SEQUENCE 572 AA; 62011 MW; 57B3BBE3A92BDEA CRC64;

Query Match 76.2%; Score 32; DB 1; Length 572;
 Best Local Similarity 83.3%; Pred. No. 36;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 6
 Db 44 WASTRH 49

RESULT 7
 TRK2_YEAST STANDARD; PRT; 889 AA.
 ID TRK2_YEAST
 AC P28584;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Potassium transport protein, low-affinity.
 GN TRK2 OR RPD2 OR YKR050W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91304419; PubMed=2072919;
 RA "Ko C.H., Gaber R.F.;
 RT "TRK1 and TRK2 encode structurally related K+ transporters in
 RT Saccharomyces cerevisiae".
 RL Mol. Cell. Biol. 11:4266-4273(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Vissers S., Urrestarazu L.A., Jauniaux J.-C.;
 RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THIS PROTEIN IS REQUIRED FOR LOW-AFFINITY POTASSIUM
 CC TRANSPORT.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M65215; AAA35172.1;
 DR EMBL: Z28275; CA82128.1;
 DR PIR: A41259; A41259.
 DR SGD: S0001758; TRK2.
 DR InterPro: IPR003445; TRKH.
 DR Pfam: PF02386; TRKH; 1.
 KW Transporth: Transmembrane; Potassium transport; Glycoprotein.
 FT TRANSMEM 52 73
 FT TRANSMEM 81 101 POTENTIAL.
 FT TRANSMEM 110 130 POTENTIAL.
 FT TRANSMEM 465 487 POTENTIAL.

FT TRANSMEM 500 521 POTENTIAL.
 FT TRANSMEM 525 545 POTENTIAL.
 FT TRANSMEM 549 569 POTENTIAL.
 FT TRANSMEM 585 605 POTENTIAL.
 FT TRANSMEM 610 630 POTENTIAL.
 FT TRANSMEM 658 678 POTENTIAL.
 FT TRANSMEM 744 764 POTENTIAL.
 FT TRANSMEM 777 797 POTENTIAL.
 FT CARBOHYD 216 216 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 606 606 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 700 700 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 889 AA; 101086 MW; 4A4148C9560276C CRC64;

Query Match 76.2%; Score 32; DB 1; Length 889;
 Best Local Similarity 83.3%; Pred. No. 57;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 6
 Db 484 WASTRH 489

RESULT 8
 VCAP_PRIVIS STANDARD; PRT; 1330 AA.
 ID VCAP_PRIVIS
 AC Q00705;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Major capsid protein (MCP) (MCP142).
 OS Pseudorabies virus (strain Indiana S) (PRV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicelloviruses.
 OX NCBI_TaxID=31522;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92024125; PubMed=1718089;
 RA Yamada S., Imada T., Watanabe W., Honda Y., Nakajima-Iijima S.,
 RA Shimizu Y., Sekikawa K.;
 RT "Nucleotide sequence and transcriptional mapping of the major capsid
 RT protein gene of pseudorabies virus".
 RL Virology 185:56-66(1991).
 CC -1- FUNCTION: MAJOR PROTEIN OF THE ICOSAEDRAL CAPSID.
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES MAJOR CAPSID PROTEIN
 CC FAMILY.
 CC PIR: A40777; VCBRS5.
 DR InterPro: IPR000912; Herpes_MCP.
 DR Pfam: PF03122; Herpes_MCP; 1.
 DR PRINTS: PR00235; HSVCAPSIDMCP.
 KW Coat protein.
 SO SEQUENCE 1330 AA; 145937 MW; 4E228145F773A522 CRC64;

Query Match 76.2%; Score 32; DB 1; Length 1330;
 Best Local Similarity 71.4%; Pred. No. 86;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 7
 Db 1205 WASTRH 1211

RESULT 9
 YGCF_ECOLI STANDARD; PRT; 223 AA.
 ID YGCF_ECOLI
 AC P55139;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)

```

DE Hypothetical protein y9cf.
CN YGCF OR B2777 OR Z4089 OR ECS3633.
OS Escherichia coli, and
05 Escherichia coli 0157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kubura S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- SIMILARITY: 70 H.INFLUENZAE H1189.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U29580; AAC69287.1; -
DR EMBL: AE000361; AAC75819.1; -
DR EMBL: AE005505; AAG57887.1; -
DR EMBL: AP002562; BAB37056.1; -
DR EcoGene: EG13034; y9cf.
DR InterPro: IPR001989; Radical_activat.
DR Pfam: PF02143; Radical_activat. 1
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 223 AA; 25029 MW; A717AAFI8F2A5D70 CRC64;

```

```

AC 084463;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CT457.
GN CT457.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/JW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Atavind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -1- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001319; AAC68057.1; -
DR InterPro: IPR002876; DUF28.
DR Pfam: PF01709; DUF28. 1.
DR ProDom: PD004323; DUF28; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 238 AA; 26563 MW; 28C001245C0D3DF6 CRC64;

```

Query Match 73.8%; Score 31; DB 1; Length 238;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 WASTRH 6
DB 7 WANTKH 12

```

RESULT 11
ID Y573.CHLPN STANDARD; PRT; 238 AA.
AC 0927Y0; Q9JQ12;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CP0176/CP0176/CP0573.
GN CP0573 OR CP0176 OR CP0573.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Utterback T., Berry K., Bass S.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CML029 from USA";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; AE001642; AAD18713.1; -
CC EMBL; AE002178; AAF38050.1; -
CC EMBL; AP002547; BAA98779.1; -
CC TIGR; CP0176; -
CC DR InterPro: IPR002876; DUF28.
CC DR Pfam: PF01709; DUF28; 1.
CC DR ProDom: PD004323; DUF28; 1.
CC DR Hypothetical protein; Complete proteome.
CC KW Hypothetical protein; Complete proteome.
CC SEQUENCE 238 AA; 26472 MW; 68E4F10F3B5FE772 CRC64;
SQ
Query Match 73.8%; Score 31; DB 1; Length 238;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 WASTRH 6
DB 7 WANTKH 12
RESULT 12
Y742_CHLMU STANDARD; PRT; 238 AA.
ID Y742_CHLMU STANDARD; PRT; 238 AA.
AC Q9PJ75;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TC0742.
OS TC0742.
GN Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mohn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uitterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; AE002342; AAF39551.1; -
CC TIGR; TC0742; -
CC DR InterPro: IPR002876; DUF28.
CC DR Pfam: PF01709; DUF28; 1.
CC DR ProDom: PD004323; DUF28; 1.
CC DR Hypothetical protein; Complete proteome.
CC KW Hypothetical protein; Complete proteome.
CC SEQUENCE 238 AA; 26486 MW; C5D5657B0909264C CRC64;
SQ
Query Match 73.8%; Score 31; DB 1; Length 238;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 WASTRH 6
DB 7 WANTKH 12
RESULT 13
Y003_MYCLE STANDARD; PRT; 251 AA.
ID Y003_MYCLE STANDARD; PRT; 251 AA.
AC Q49645;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 26.6 kDa protein ML0475.
GN ML0475 OR U1177B OR B1177_C2_181.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CC NCB1_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Gardner T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus";
RL Nature 403:1007-1011(2001).
CC -1- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; U00011; AAA17094.1; -
CC EMBL; AL583918; CAC29983.1; -
CC Leproma; ML0475; -
CC DR InterPro: IPR002876; DUF28.
CC DR Pfam: PF01709; DUF28; 1.
CC DR ProDom: PD004323; DUF28; 1.
CC DR Hypothetical protein; Complete proteome.
CC SEQUENCE 251 AA; 26614 MW; 8584C11AE1F02E6C CRC64;
SQ

FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 445 AA; 50495 MW; 31CA169CB099F194 CRC64;

Query Match 73.8%; Score 31; DB 1; Length 445;
Best Local Similarity 83.3%; Pred. No. 44;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 6
11111

Db 269 WAVTRH 274

Search completed: August 20, 2002, 11:41:21
Job time: 400 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:40:53 ; Search time 43.57 Seconds
(without alignments)
27.794 Million cell updates/sec

Title: US-09-824-286-4_COPY_50_56
Perfect score: 42
Sequence: 1 WASTRHT 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	387	2 Q9R6M6	Q9R6M6 agrobacteri
2	34	81.0	199	10 Q9LVEF8	Q9LVEF8 arabidopsi
3	34	81.0	212	5 Q62508	Q62508 caenorhabd
4	34	81.0	435	10 Q9FMJ7	Q9FMJ7 arabidopsi
5	34	81.0	605	3 Q9P564	Q9P564 neurospora
6	34	81.0	743	10 Q9LW44	Q9LW44 lycopersico
7	34	81.0	746	10 Q9LW43	Q9LW43 lycopersico
8	33	78.6	206	16 Q9LW43	Q9LW43 lycopersico
9	33	78.6	313	5 Q95QL5	Q95QL5 pseudomonas
10	33	78.6	347	16 Q9L5J7	Q9L5J7 pseudomonas
11	33	78.6	650	5 Q95U14	Q95U14 drosophila
12	32	76.2	132	4 Q9BUN6	Q9BUN6 homo sapien
13	32	76.2	142	16 Q9PC08	Q9PC08 xylella fas
14	32	76.2	222	2 Q9X5R1	Q9X5R1 zymomonas m
15	32	76.2	263	2 Q84969	Q84969 rhodococcus
16	32	76.2	268	16 Q06172	Q06172 mycobacteri

17	32	76.2	268	16 Q07709	Q07709 mycobacteri
18	32	76.2	337	4 Q9NZ76	Q9NZ76 homo sapien
19	32	76.2	345	4 Q9NZK9	Q9NZK9 homo sapien
20	32	76.2	434	16 Q92A01	Q92A01 listeria in
21	32	76.2	445	2 Q93J94	Q93J94 streptomyc
22	32	76.2	446	1 Q9V2V8	Q9V2V8 thermoprote
23	32	76.2	462	2 Q9FBI5	Q9FBI5 streptomyc
24	32	76.2	503	12 Q9J303	Q9J303 ectocarpus
25	32	76.2	672	5 Q9NK74	Q9NK74 drosophila
26	32	76.2	725	16 Q9RS62	Q9RS62 deinococcus
27	32	76.2	738	10 Q48929	Q48929 nicotiana t
28	32	76.2	745	16 Q9RT19	Q9RT19 deinococcus
29	32	76.2	818	5 Q9VJN0	Q9VJN0 drosophila
30	32	76.2	824	10 Q64776	Q64776 arabidopsi
31	32	76.2	1330	12 Q04534	Q04534 pseudorabie
32	32	76.2	1385	12 Q65565	Q65565 bovine herp
33	32	76.2	1468	5 Q9V854	Q9V854 drosophila
34	32	76.2	2334	12 Q96725	Q96725 european br
35	31	73.8	127	4 Q96NHL	Q96NHL homo sapien
36	31	73.8	196	10 Q9ZPT2	Q9ZPT2 arabidopsi
37	31	73.8	198	2 Q9F1M0	Q9F1M0 streptomyc
38	31	73.8	237	2 Q9RJT0	Q9RJT0 streptomyc
39	31	73.8	250	2 Q9L288	Q9L288 streptomyc
40	31	73.8	251	2 Q9AE12	Q9AE12 corneabacte
41	31	73.8	288	2 Q51543	Q51543 pseudomonas
42	31	73.8	394	2 Q914X0	Q914X0 streptomyc
43	31	73.8	429	5 Q9NB72	Q9NB72 trypanosoma
44	31	73.8	448	2 Q93076	Q93076 rhodococcus
45	31	73.8	448	10 Q9XHG1	Q9XHG1 graciilaria

ALIGNMENTS

RESULT	ID	Q9R6M6	PRELIMINARY:	PRT:	387 AA.
Q9R6M6	Q9R6M6	Q9R6M6	Q9R6M6	Q9R6M6	Q9R6M6
AC	Q9R6M6	Q9R6M6	Q9R6M6	Q9R6M6	Q9R6M6
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	TIOF33	PROTEIN.			
GN	TIOF33				
OS	Agrobacterium tumefaciens.				
OC	Plasmid pTi-SAKURA.				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;				
OC	Rhizobiaceae; Rhizobium.				
OX	NCBI_TaxID=358;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MAFF301001;				
RX	MEDLINE=20184752; PubMed=10721727;				
RA	Suzuki K., Hattori Y., Uraji M., Ohta N., Iwata K., Murata K.,				
RA	Katoh A., Yoshida K.;				
RT	"Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.";				
RL	Gene 242:331-336(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MAFF301001;				
RX	MEDLINE=98193120; PubMed=9524202;				
RA	Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;				
RT	"Novel structural difference between nopaline- and octopine- type ttrb				
RT	gene: construction of genetic and physical map and sequencing of				
RT	ttr/traI and rep gene clusters of a new Ti plasmid pTi-SAKURA.";				
RL	Biochim. Biophys. Acta 1396:1-7(1998).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MAFF301001;				
RA	Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;				
RT	"Genome structure of pTi-SAKURA(T): Strategy for DNA sequencing of a				
RT	Japanese cherry-Ti plasmid.";				
RL	Nucleic Acids Symp. Ser. 37:159-160(1998).				
RN	[4]				

RP SEQUENCE FROM N.A.
RC STRAIN-MAFF301001;
RA Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pri-SAKURA (III): Characteristics of T-DNA.";
RL Nucleic Acids Symp. Ser. 39:185-186(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-MAFF301001;
RA Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;
RT "Genome structure of pri-SAKURA (IV): Characteristics of tra region.";
RL Nucleic Acids Symp. Ser. 39:187-188(1998).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN-MAFF301001;
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pri-SAKURA (V): Complete nucleotide sequence of
RT plasmid pri-SAKURA's vir region in Agrobacterium tumefaciens.";
RL Nucleic Acids Symp. Ser. 39:265-266(1998).
DR EMBL: AB016260; BAA87658.1; -.
KW Plasmid.
SQ SEQUENCE 387 AA; 43706 MW; 67AE002F5A6684F6 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 387;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6
DB 120 WASTRH 125

RESULT 2
O9LVF8 PRELIMINARY; PRT; 199 AA.
AC O9LVF8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 3, P1 CLONE: M1L23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and YAC
RT clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL: AB019232; BAB02342.1; -.
SQ SEQUENCE 199 AA; 22873 MW; 2ABE82C5727F7AE9 CRC64;

Query Match 81.0%; Score 34; DB 10; Length 199;
Best Local Similarity 83.3%; Pred. No. 24;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6
DB 107 WASTRH 112

RESULT 3

us-09-824-286-4_copy_50_56.rspt

O62508 PRELIMINARY; PRT; 212 AA.
AC O62508;
ID O62508;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ZK384.2 PROTEIN.
GN ZK384.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nemata; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Ainscough R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z82092; CAB05011.1; -.
DR InterPro: IPR001283; SCP.
DR Pfam: PF00188; SCP; 2.
DR PRINTS: PR00837; V5TPXLIKE.
DR SMART: SM00198; SCP; 1.
SQ SEQUENCE 212 AA; 24122 MW; B25A6D727EAC3B7 CRC64;

Query Match 81.0%; Score 34; DB 5; Length 212;
Best Local Similarity 83.3%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6
DB 145 WASTRH 150

RESULT 4
O9FMJ7 PRELIMINARY; PRT; 435 AA.
AC O9FMJ7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE SIMILARITY TO ZINC FINGER PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., MiyaJima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned P1 clones.";
RL DNA Res. 4:401-414(1997).
DR EMBL: AB008265; BAB10568.1; -.
DR InterPro: IPR000571; ZF-CCCH.
DR Pfam: PF00642; ZF-CCCH; 5.
DR SMART: SM00356; ZNF_C3H1; 5.
SQ SEQUENCE 435 AA; 48606 MW; 8BACBB780A03C6F2 CRC64;

Query Match 81.0%; Score 34; DB 10; Length 435;
Best Local Similarity 83.3%; Pred. No. 55;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6
1:||||
Db 254 WSTRH 259

RESULT 5
ID 09P564 PRELIMINARY; PRT; 605 AA.
AC 09P564;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE RELATED TO BETA TRANSDUCIN-LIKE PROTEIN.
GN B9J10.170.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL256324; CAB92024.1; -
SQ SEQUENCE 605 AA; 69676 MW; 1ED03092C7A894A3 CRC64;

Query Match 81.0%; Score 34; DB 3; Length 605;
Best Local Similarity 71.4%; Pred. No. 78;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 WASTRH 7
1:||||
Db 201 WASTRH 207

RESULT 6
ID 09LMA4 PRELIMINARY; PRT; 743 AA.
AC 09LMA4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SUBTILISIN-LIKE PROTEASE.
GN P69E.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RX MEDLINE=20098714; PubMed=10631250;
RA Jorda L., Conejero V., Vera P.;
RT "Characterization of two differentially regulated genes (P69E and P69F) encoding new members of the subtilisin-like protease clan from tomato plants.";
RL Plant Physiol. 122:67-76(2000).
DR EMBL: Y18931; CAB67119.1; -
DR HSSP: Q99405; 1MPT.
DR MEROPS: S08.006; -
DR InterPro: IPR003137; PA.
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.

DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 743 AA; 79134 MW; 63EE2F2B2510BCA CRC64;

Query Match 81.0%; Score 34; DB 10; Length 743;
Best Local Similarity 71.4%; Pred. No. 97;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 WASTRH 7
1:||||
Db 727 WSTRHS 733

RESULT 7
ID 09LMA3 PRELIMINARY; PRT; 746 AA.
AC 09LMA3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SUBTILISIN-LIKE PROTEASE.
GN P69F.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RX MEDLINE=20098714; PubMed=10631250;
RA Jorda L., Conejero V., Vera P.;
RT "Characterization of two differentially regulated genes (P69E and P69F) encoding new members of the subtilisin-like protease clan from tomato plants.";
RL Plant Physiol. 122:67-76(2000).
DR EMBL: Y18932; CAB67120.1; -
DR HSSP: P29599; 1ST3.
DR MEROPS: S08.006; -
DR InterPro: IPR003137; PA.
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 746 AA; 79062 MW; F2B9D6B7EA0E94AB CRC64;

Query Match 81.0%; Score 34; DB 10; Length 746;
Best Local Similarity 71.4%; Pred. No. 98;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 WASTRH 7
1:||||
Db 730 WSTRHS 736

RESULT 8
ID 09I681 PRELIMINARY; PRT; 206 AA.
AC 09I681;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE PROBABLE TRANSCRIPTIONAL REGULATOR.
GN PA0436.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AE004481; AAC03825.1; -
DR InterPro: IPR001647; HTH_Tetr.
DR Pfam: PF00440; tetr. 1.
DR PRINTS; PR00455; HTHTETR.
KW Complete proteome: DNA-binding; Transcription regulation.
SQ SEQUENCE 206 AA; 23718 MW; 365E298E435FF1B CRC64;

Query Match 78.6%; Score 33; DB 16; Length 206;
Best Local Similarity 83.3%; Pred. No. 40;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WASTRH 6
Db 160 WASTQH 165

RESULT 9
Q950L5 PRELIMINARY; PRT; 313 AA.
AC Q950L5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHEICAL 36.3 KDA PROTEIN.
GN F10E9.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Anderson K.;
RT "The sequence of C. elegans cosmid F10E9.";
RL Submitted (Feb-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (Nov-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; L10986; AAK93846.1; -
KW Hypothetical protein.
SQ SEQUENCE 313 AA; 36253 MW; 9D6C3002C30F9AF3 CRC64;

Query Match 78.6%; Score 33; DB 5; Length 313;
Best Local Similarity 83.3%; Pred. No. 62;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

Qy 1 WASTRH 6
Db 164 WASTRH 169

RESULT 10
Q915J7 PRELIMINARY; PRT; 347 AA.
AC Q915J7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE HYPOTHEICAL PROTEIN PA0732.
GN PA0732.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004508; AAC04121.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 347 AA; 37161 MW; 264E4DC064290CC3 CRC64;

Query Match 78.6%; Score 33; DB 16; Length 347;
Best Local Similarity 85.7%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WASTRH 7
Db 318 WASTRDT 324

RESULT 11
Q950I4 PRELIMINARY; PRT; 650 AA.
AC Q950I4;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GH14380P.
GN CG12239.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN=Y, CN BW SP;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunco J., Pacleb J., Paragas V., Park S., Phoumenavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY058378; AAL13607.1; -
SQ SEQUENCE 650 AA; 80172 MW; 106796301C835D67 CRC64;
```

Query Match 78.6%, Score 33; DB 5; Length 650;
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6
 |||||
 Db 22 WASSRH 27

RESULT 12

O9BUN6 PRELIMINARY; PRT; 132 AA.

AC O9BUN6;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE SIMILAR TO PROGRAMMED CELL DEATH 9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-EYE, RETINOBLASTOMA;
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC002460; AAH02460.1; -
 SQ SEQUENCE 132 AA; 14563 MW; 749ED93CF43D9A34 CRC64;

Query Match 76.2%; Score 32; DB 4; Length 132;
 Best Local Similarity 71.4%; Pred. No. 39;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 7
 |||||
 Db 85 WASSRH 91

RESULT 13

O9PG08 PRELIMINARY; PRT; 142 AA.

AC O9PG08;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHEICAL PROTEIN XF0240.
 GN XF0240.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Rehnard F.C., Arruda P., Abreu F.A., Acencio M.,
 Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
 Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
 Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 Facinanci A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitaajima J.P.,
 Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,
 Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 Menck C.F.M., Miraca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 Nham A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,

RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,
 Ra Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 Ra Queiro R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 DR EMBL: AE003878; AAF83053.1; -
 DR InterPro: IPR002577; DUF24.
 DR Pfam: PF01638; DUF24; 1.
 DR Prodom: PD004032; DUF24; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 142 AA; 15878 MW; CAD3A2B6EDB13E31 CRC64;

Query Match 76.2%; Score 32; DB 16; Length 142;
 Best Local Similarity 83.3%; Pred. No. 43;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 6
 |||||
 Db 113 WASTRH 118

RESULT 14

O9X5F1 PRELIMINARY; PRT; 222 AA.

AC O9X5F1;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHEICAL 25.0 KDA PROTEIN.
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
 OC Zymomonas.
 OX NCBI_TaxID=542;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ZM4;
 RA Lee H.J., Kang H.S.;
 RT "Sequence analysis of 43D2 fosmid clone of Zymomonas mobilis ZM4.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF124757; AAD29658.1; -
 DR InterPro: IPR001441; UPP_synth.
 DR Pfam: PF01255; UPP_synthetase; 1.
 DR Prodom: PD003461; UPP_synth; 1.
 DR PROSITE: PS01066; UPP_SYNTHETASE; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 222 AA; 24977 MW; EA9723DCB30F42FB CRC64;

Query Match 76.2%; Score 32; DB 2; Length 222;
 Best Local Similarity 83.3%; Pred. No. 69;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 6
 |||||
 Db 7 WASSRH 12

RESULT 15

O84989 PRELIMINARY; PRT; 263 AA.

AC O84989;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PUTATIVE TRANSPOSASE (FRAGMENT).
 OS Rhodococcus opacus (Nocardia opaca).

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_taxid=37919;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1CP;
RX MEDLINE=98324954; PubMed=9657989;
RA Seibert V., Kourbatova E.M., Golovleva L.A., Schlomann M.;
RT "Characterization of the maleylacetate reductase MacA of Rhodococcus
RT opacus 1CP and evidence for the presence of an isofunctional enzyme.";
RL J. Bacteriol. 180:3503-3508(1998).
DR EMBL; AF030176; AAC38799.1; -;
DR InterPro: IPR002525; Transposase_9.
DR Pfam; PF01548; Transposase_9; 1.
FT NON_TER 263 263
SQ SEQUENCE 263 AA; 28327 MW; 64DBAD92CB4317D9 CRC64;

Query Match 76.2%; Score 32; DB 2; Length 263;
Best Local Similarity 71.4%; Pred. No. 82;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 WASTRHT 7
1 ||||
DB 151 WERTRHT 157

Search completed: August 20, 2002, 11:40:54
Job time: 433 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:15:11 ; Search time 94.22 Seconds
(without alignments)
11.789 Million cell updates/sec

Title: US-09-824-286-4_COPY_95_104
Perfect score: 61
Sequence: 1 PWTFCGCTKL 10

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	19	21	AA801631
2	61	100.0	20	14	AA833254
3	61	100.0	104	13	AA830010
4	61	100.0	106	19	AA831647
5	61	100.0	107	12	AA811985
6	61	100.0	107	13	AA825411
7	61	100.0	107	14	AA838614
8	61	100.0	107	14	AA81283
9	61	100.0	107	19	AA866098
10	61	100.0	107	19	AA849812
11	61	100.0	107	19	AA858515

12	61	100.0	107	19	AA858505	H65 protein sequen
13	61	100.0	107	19	AA858478	Humanised anti-CD5
14	61	100.0	108	13	AA821292	Murine VL kappa gr
15	61	100.0	108	13	AA821310	Light chain of Mf
16	61	100.0	108	16	AA879160	Human IgE receptor
17	61	100.0	108	17	AA800834	Variable light cha
18	61	100.0	108	17	AA804177	Variant variable 1
19	61	100.0	108	17	AA800828	Variable light cha
20	61	100.0	108	18	AA816620	Anti-human FasL an
21	61	100.0	108	18	AA827360	Light chain variab
22	61	100.0	108	18	AA827355	Light chain variab
23	61	100.0	108	19	AA807620	Anti-VEGF murine a
24	61	100.0	108	22	AA807484	Synthetic antibody
25	61	100.0	108	22	AA807486	Synthetic antibody
26	61	100.0	108	22	AA807488	Synthetic antibody
27	61	100.0	108	22	AA807480	Synthetic antibody
28	61	100.0	108	22	AA807482	Synthetic antibody
29	61	100.0	108	22	AA807492	Synthetic antibody
30	61	100.0	108	22	AA807496	Synthetic antibody
31	61	100.0	108	22	AA807517	Antibody scFv8 11g
32	61	100.0	108	22	AA881980	Ganglioside GD3 sp
33	61	100.0	109	16	AA879858	Anti-EGFR antibody
34	61	100.0	109	18	AA818271	Prp 37 light chain
35	61	100.0	109	22	AA85910	Anti-Prp antibody
36	61	100.0	109	22	AA85862	Light chain variab
37	61	100.0	110	13	AA830012	Light chain variab
38	61	100.0	110	15	AA860810	Murine anti-HIV mu5
39	61	100.0	111	15	AA855123	Anti HIV antibody
40	61	100.0	111	15	AA860302	Murine M195 antibo
41	61	100.0	111	22	AA869669	VI region from an
42	61	100.0	112	16	AA880272	Variable light cha
43	61	100.0	112	17	AA800830	Murine 5D12 monocl
44	61	100.0	112	18	AA823024	Humanized 5D12 mon
45	61	100.0	112	18	AA823026	

ALIGNMENTS

RESULT 1	
AA801631	standard; Peptide: 19 AA.
ID	AA801631;
XX	
AC	AA801631;
XX	
DT	07-DEC-2000 (first entry)
XX	
DE	Chimaeric immunoglobulin ligation sequence #3.
XX	
KW	Immunoglobulin: variable region; Light chain; L chain; cancer;
XX	humanised antibody.
OS	Chimeric - Mus sp.
OS	Chimeric - Homo sapiens.
XX	
PN	EP1013761-A2.
XX	
PD	28-JUN-2000.
XX	
PF	18-SEP-1992; 99EP-0124345.
XX	
PR	18-SEP-1991; 91JP-0238375.
PR	18-SEP-1992; 92EP-0116026.
XX	
PA	(KYOW) KYOWA HAKKO KOGYO KK.
XX	
PI	Shitara K, Hanai N, Hasegawa M, Miyaji H, Kuwana Y;
XX	
DR	WPI: 2000-402204/35.
DR	N-PSDB: AAA51007.
XX	
PT	New humanized chimera antibody KM-871 useful for treating cancer,
	comprises variable region of mouse monoclonal antibody, reactive with

PT ganglioside and human antibody constant region -
 XX
 PS Example 2; Page 29; 65pp; English.
 CC
 CC The present sequence is a ligation sequence used to join the protein
 CC sequence for a murine immunoglobulin light chain variable region from
 CC plasmid KM-641 with a human light chain antibody protein sequence. This
 CC was used in an expression vector to produce humanised chimaeric
 CC antibodies, which can be used to treat cancer. Humanised chimaeric
 CC antibodies are more effective than mouse antibodies as they do not
 CC provoke a reaction in the human and side effects, such as the formation
 CC of anti-mouse immunoglobulin antibody and the rapid half-life of the
 CC immunoglobulins, do not occur.
 XX
 SQ Sequence 19 AA;

Query Match 100.0%; Score 61; DB 21; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.0068;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTFGGGTKL 10
 |||||
 Db 2 PWTFGGGTKL 11

RESULT 2
 AAR33254
 ID AAR33254 standard; Protein: 20 AA.
 XX
 AC AAR33254;

XX 12-JUL-1993 (first entry)
 XX
 DE PKM641A2 derived L chain linker.

XX Linker; humanised; chimeric; antibody; expression vector; heavy;
 KW light; chain; variable region.
 XX

XX Synthetic.

XX EP53199-A.

XX 24-MAR-1993.

XX 18-SEP-1992; 92EP-0116026.

XX 18-SEP-1991; 91JP-0238375.

XX (KYOW) KYOWA HAKKO KOGYO CO LTD.

XX Hanai N, Hasegawa M, Kuwana Y, Miyaji H, Shitara K;

XX WPI: 1993-095510/12.

XX N-PSDB; AAQ37055.

XX Humanised chimeric antibody prodn. against ganglioside GD3 - for
 PT treating cancers, such as melanoma, neuroblastoma, etc.

XX Claim 15; Page 31; 63pp; English.

XX The sequences given in AAR33253-54 are linkers which were used in the
 CC construction of humanised chimeric antibodies. The linkers are
 CC used in construction of heavy and light chain variable region
 CC antibodies respectively.

XX Sequence 20 AA;

Query Match 100.0%; Score 61; DB 14; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.0071;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTFGGGTKL 10
 |||||
 Db 2 PWTFGGGTKL 11

RESULT 3
 AAR30010
 ID AAR30010 standard; Protein: 104 AA.
 XX
 AC AAR30010;

XX 29-APR-1993 (first entry)

XX Light chain variable region of anti-p24 antibody.

XX MAB: monoclonal antibody; AIDS; HIV; infection; disease; immunoassay;
 KW p24; antibody-dependent cell-mediated cytotoxicity; macrophages;
 KW lymphoid cells.

XX Chimaeric; Homo sapiens; Mus musculus.

XX EP519866-A.

XX 23-DEC-1992.

XX 10-JUN-1993; 93EP-0810445.

XX 18-JUN-1991; 91EP-0810468.

XX (CIBA) CIBA GEIGY AG.

XX MESTAN J, LAZDINS JK, WOODS-COOK KA, HARDMAN N, HOCHKEPPEL H;

XX WPI: 1992-426048/52.

XX Monoclonal antibody specific for HIV core protein p24 - mediate
 PT antibody-dependent cell-mediated cytotoxicity, for treating,
 PT preventing and diagnosing HIV infection.

XX Example 4; Page 22; 44pp; English.

XX This sequence represents the light chain variable region from
 CC anti-p24 antibody from hybridoma 25-57-1. The MAB is used to
 CC prevent progression of AIDS and to treat HIV infections, and
 CC also as an immunoassay reagent for qualitative or quantitative
 CC analysis of p24 for diagnostic purposes.

XX Sequence 104 AA;

Query Match 100.0%; Score 61; DB 13; Length 104;
 Best Local Similarity 100.0%; Pred. No. 0.034;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTFGGGTKL 10
 |||||
 Db 93 PWTFGGGTKL 102

RESULT 4
 AAW31647
 ID AAW31647 standard; Protein: 106 AA.
 XX
 AC AAW31647;

XX 21-MAY-1998 (first entry)

XX Monoclonal antibody CP.B8 light chain variable region.

XX Cytokine receptor; gamma common chain; gc chain; human;
 KW blocking agent; monoclonal antibody; CP.B8; immunological disease;
 KW myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
 KW insulin-dependent diabetes; inflammatory bowel disease;

KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;
 KW graft versus host disease; psoriasis; immunosuppressive; therapy;
 KW complementarity determining region; CDR.
 OS Mus musculus.
 XX
 XX
 FH Key Location/Qualifiers
 FT Region 24..34
 FT /note= "CDR1"
 FT Region 50..56
 FT /note= "CDR2"
 FT Region 89..97
 FT /note= "CDR3"
 PN MO9743416-A1.
 XX
 XX
 PD 20-NOV-1997.
 XX
 XX
 PF 09-MAY-1997; 97MO-US07870.
 XX
 XX
 PR 10-MAY-1996; 96US-0017466.
 XX
 XX
 PA (BIOJ) BIOGEN INC.
 XX
 PI Benjamin CD, Burkly LC, Hession C, Whitty A;
 XX
 DR WPI: 1998-008885/01.
 DR N-PSDB; AAT97440.
 XX
 XX
 PT Blocking agents of the gamma common chain of cytokine receptors -
 PT particularly monoclonal antibodies, used to induce T cell anergy for
 PT treatment of immunological diseases
 XX
 PS Claim 22; Page 81; 111pp; English.
 XX
 CC This polypeptide comprises the light chain variable region (VL) of
 CC monoclonal antibody (MAb) C9.B8, which is produced by a hybridoma
 CC deposited as ATCC 12107, and which is specific for the gamma
 CC constant (gc) chain (see AAW31646) of human cytokine receptors. The
 CC invention provides compositions and methods for inhibiting cytokine
 CC signalling using gc chain blocking agents for the treatment of
 CC immunological diseases such as myasthenia gravis, rheumatoid
 CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,
 CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,
 CC allergy, asthma, parasitic infection, graft vs. host disease or
 CC psoriasis. Preferred gc blocking agents include MAb CP.B8, its Fab
 CC fragment and an antibody having a light chain variable region
 CC CDR selected from those of CP.B8 VL or a heavy chain variable
 CC region CDR selected from those of CP.B8 VH (see AAW31648).
 CC
 XX
 SQ Sequence 106 AA:
 Query Match 100.0%; Score 61; DB 19; Length 106;
 Best Local Similarity 100.0%; Pred. No. 0.034;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PWTFGGGTKL 10
 Db 95 PWTFGGGTKL 104
 RESULT 5
 ID AAR11985 standard; Protein; 107 AA.
 AC AAR11985;
 XX
 DT 31-JUL-1991 (first entry)
 DE Anti-placental alkaline phosphatase Vx.
 XX
 KW Placental alkaline phosphatase; antibody; cancer; light chain;
 XX

KW PLAP; CDR.
 XX
 OS Mus musculus.
 XX
 XX
 FH Key Location/Qualifiers
 FT Region 24..34
 FT /label= CDR1
 FT Region 50..56
 FT /label= CDR2
 FT Region 89..97
 FT /label= CDR3
 PN EP429242-A.
 XX
 XX
 PD 29-MAY-1991.
 XX
 XX
 PF 14-NOV-1990; 90EP-0312407.
 XX
 XX
 PR 07-SEP-1990; 90GB-0019552.
 PR 17-NOV-1989; 89GB-0026045.
 XX
 XX
 PA (UNIL) UNILEVER PLC.
 XX
 PI Verhoeven ME;
 XX
 DR WPI: 1991-157662/22.
 DR N-PSDB; AAQ11846.
 XX
 XX
 PT Synthetic specific binder for human placental alkaline
 PT phosphatase - for treating and diagnosing cancers e.g. breast,
 PT ovarian and colon cancers and pleural effusions
 XX
 PS Disclosure; Fig 2; 28pp; English.
 XX
 CC The murine anti-PLAP Vx gene contains 3 CDRs which are used to reshape
 CC the human anti-PLAP Vx gene (AAQ11848). The murine anti-PLAP Vx gene
 CC (AAQ11845) CDRs were used to reshape the human anti-PLAP gene Vx
 CC (AAQ11847). The specific binder for human PLAP contains at least
 CC one of the former 3 CDRs and/or at least one of the latter CDRs.
 CC The produced antibodies are more readily tolerated when administered
 CC to a human patient. Antibody reagents can be used to identify, e.g.
 CC by serum testing or imaging, and/or to treat PLAP-producing cancers.
 CC Such cancers can occur as, e.g. breast cancer, ovarian cancer and
 CC colon cancer or can manifest themselves as liquids such as pleural
 CC effusions.
 CC
 XX
 SQ Sequence 107 AA:
 Query Match 100.0%; Score 61; DB 12; Length 107;
 Best Local Similarity 100.0%; Pred. No. 0.035;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PWTFGGGTKL 10
 Db 95 PWTFGGGTKL 104
 RESULT 6
 ID AAR25411 standard; Protein; 107 AA.
 AC AAR25411;
 XX
 DT 11-JAN-1993 (first entry)
 DE Light chain variable domain of MAb CB6 to TNFalpha epitope.
 XX
 KW monoclonal antibody; humanised; chimeric; CDR; neoplastic therapy;
 KW shock; antilymphocyte therapy; endotoxin shock; septic shock;
 KW sepsis; cardiovascular shock; tumour necrosis factor alpha;
 KW multi-organ failure.
 XX

CS Chimeric.
 XX
 PN WO9211383-A.
 XX
 PD 09-JUL-1992.
 XX
 PE 20-DEC-1991; 91WO-GB02300.
 XX
 PR 21-DEC-1990; 90WO-GB02017.
 PR 03-MAY-1991; 91GB-0009645.
 XX
 PA (CLLT) CELLTECH LTD.
 XX
 PI Adair JR, Althwal DS, Bodmer MW, Emtage JS;
 XX
 DR WPI: 1992-250102/30.
 XX
 PT Recombinant antibody specific for human TNF-alpha - for treatment
 PT of shock and immuno:regulatory and inflammatory disorders, also
 PT used in diagnosis
 PS
 PS Claim 2; Fig 3; 57pp; English.
 XX
 CC This sequence is that of the light chain variable domain of a chimeric
 CC antibody to an epitope of human tumour necrosis factor. The CDR-grafting
 CC of this light chain has been based on the REI light chain. A gene
 CC was built which coded for amino acid residues 23, 24, 48, 49, 71,
 CC 73, 75, 88 [GH341(8)] as murine residues. It was then co-expressed
 CC with g1221, the minimum grafted light chain (CDRs only).
 CC
 XX
 SQ Sequence 107 AA;
 OY 1 PWTFGGGTRL 10
 DB 95 pwtfgggtrl 104
 RESULT 7
 AAR38614
 ID AAR38614 standard; peptide: 107 AA.
 XX
 AC AAR38614;
 XX
 DE 28-OCT-1993 (first entry)
 XX
 DE Low-risk modified (prop) light chain of H65 Mab.
 XX
 KW Antibody: variable domain; light; L; heavy; H; consensus;
 KW affinity: antigen; immunogenicity: humanisation; framework;
 KW monoclonal antibody; Mab; hkl; h3.
 XX
 OS Mus sp.
 OS
 PN WO9311794-A.
 XX
 PD 24-JUN-1993.
 XX
 PE 14-DEC-1992; 92WO-US10906.
 XX
 PR 13-DEC-1991; 91US-0808464.
 XX
 PA (XOMA) XOMA CORP.
 XX
 PI Fishwild DM, Kohn FR, Little RG, Studnicka GM;
 XX
 DR WPI: 1993-213827/26.
 XX
 PT Antibodies prepn. used for treatment of auto-immune diseases - by

PT replacement of critical residues to reduce immunogenicity but
 PT retain binding affinity, etc.
 XX
 PS Example 1; Fig 6A; 160pp; English.
 XX
 CC Modified antibody variable domains were prepd. by identifying
 CC low-risk residues in a mouse monoclonal antibody variable domain,
 CC designated H65, which may be modified without diminishing the
 CC native affinity of the domain for antigen while still reducing
 CC its immunogenicity w.r.t. humans. The light and heavy chains of
 CC the variable domain of H65 were determined to most closely resemble
 CC the consensus sequences of subgroup 1 (hkl) of the human kappa
 CC chains and subgroup 3 (h3) of the human heavy chains, respectively.
 CC The H65 V/J-segments of the light and heavy chain sequences are given
 CC in AAR38614 and AAR38615 respectively.
 XX
 SQ Sequence 107 AA;
 OY 1 PWTFGGGTRL 10
 DB 95 pwtfgggtrl 104
 RESULT 8
 AAR41283
 ID AAR41283 standard; peptide: 107 AA.
 XX
 AC AAR41283;
 XX
 DE 28-OCT-1993 (first entry)
 XX
 DE Moderate-risk modified (prop) light chain of H65 Mab.
 XX
 KW Antibody: variable domain; light; L; heavy; H; consensus;
 KW affinity: antigen; immunogenicity: humanisation; framework;
 KW monoclonal antibody; Mab; hkl; h3.
 XX
 OS Mus sp.
 OS
 PN WO9311794-A.
 XX
 PD 24-JUN-1993.
 XX
 PE 14-DEC-1992; 92WO-US10906.
 XX
 PR 13-DEC-1991; 91US-0808464.
 XX
 PA (XOMA) XOMA CORP.
 XX
 PI Fishwild DM, Kohn FR, Little RG, Studnicka GM;
 XX
 DR WPI: 1993-213827/26.
 XX
 PT Antibodies prepn. used for treatment of auto-immune diseases - by
 PT replacement of critical residues to reduce immunogenicity but
 PT retain binding affinity, etc.
 XX
 PS Example 6; Fig 16A; 160pp; English.
 XX
 CC Modified antibody variable domains were prepd. by identifying
 CC moderate-risk residues in a mouse monoclonal antibody variable domain,
 CC designated H65, which may be modified without diminishing the
 CC native affinity of the domain for antigen while still reducing
 CC its immunogenicity w.r.t. humans. The light and heavy chains of
 CC the variable domain of H65 were determined to most closely resemble
 CC the consensus sequences of subgroup 1 (hkl) of the human kappa
 CC chains and subgroup 3 (h3) of the human heavy chains, respectively.
 CC The H65 V/J-segments of the light and heavy chain sequences are given

CC in AAR42183 and AAR42184 respectively.
 XX
 SQ Sequence 107 AA;

Query Match 100.0%; Score 61; DB 14; Length 107;
 Best Local Similarity 100.0%; Pred. No. 0.035;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTXL 10
 DB 95 pwtfgggtkl 104

RESULT 9
 AAM66098
 ID AAM66098 standard; Protein; 107 AA.
 XX
 AC AAM66098;

DE 10-DEC-1998 (first entry)

XX anti-CD22 monoclonal antibody light chain variable region.

XX anti-CD22 monoclonal antibody light chain variable region; VL;

KW Pseudomonas exotoxin; variable heavy chain; VH; variable light chain;

KM malignant B-cell; Immunodiagnosis; RFB4 IgG.

XX Mammalia.

XX WO9841641-A1.

XX 24-SEP-1998.

XX 19-MAR-1998; 98WO-US05453;

XX 20-MAR-1997; 97US-0041437.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Fitzgerald D, Kreitman R, Mansfield E, Pastan I;

XX WPI; 1998-521227/44.

XX N-PSDB; AAV07641.

XX Recombinant anti-CD22 antibodies and immuno-conjugates - of

XX or a label; for inhibiting malignant B-cells

XX Claim 6; Fig 1; 71pp; English.

XX The invention claims for a recombinant immunconjugate comprising

XX of a therapeutic agent (e.g. Pseudomonas exotoxin) or a detectable

XX label peptide bonded to a recombinant anti-CD22 antibody (RFB4 IgG)

XX having a variable heavy (VH; AAM66099) chain with a cysteine residue

XX at amino acid 44 and the present variable light (VL) chain with a

XX cysteine residue at amino acid 100. The immunconjugate is claimed

XX to inhibit the growth of malignant B-cells in vivo, such as rodent,

XX canine or primate B-cells. The anti-CD22 antibody is claimed useful

XX for detecting CD22 protein in a sample or in vivo in a mammal, and

XX can be used in diagnostic kits.

XX Sequence 107 AA;

Query Match 100.0%; Score 61; DB 19; Length 107;
 Best Local Similarity 100.0%; Pred. No. 0.035;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTXL 10
 DB 95 pwtfgggtkl 104

RESULT 10
 AAM49812
 ID AAM49812 standard; Protein; 107 AA.
 XX
 AC AAM49812;

DE 24-SEP-1998 (first entry)

XX Amino acid sequence of the mouse antibody C4G1 mature light chain.

XX Light chain; humanised; immunoglobulin; Ig; mouse C4G1; antibody;

XX inhibition; antigen; cardiovascular disease; thromboembolic disorder;

XX cancer; acute myocardial infarction; unstable angina; stroke;

XX transient ischemic episode; pulmonary embolism; deep vein thrombosis;

XX extracorporeal cardiopulmonary circulation.

XX Mus sp.

XX Key

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

Query Match 100.0%; Score 61; DB 19; Length 107;
 Best Local Similarity 100.0%; Pred. No. 0.035;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTXL 10
 DB 95 pwtfgggtkl 104

Claim 1; Fig 5A; 35pp; English.

XX This is the amino acid sequence of the humanised antibody C4G1 light

XX chain, used in the method of the invention involving the creation

XX of a humanised immunoglobulin (Ig) derived from the mouse C4G1 antibody.

XX The humanised Ig is capable of binding to GPIIb/IIIa and inhibiting

XX platelet aggregation and also the releasing reaction of platelets. The

XX Ig can be used for treating cardiovascular diseases and thromboembolic

XX disorders, e.g. acute myocardial infarction, unstable angina, stroke,

XX transient ischemic episodes, deep vein thrombosis and pulmonary embolism,

XX extracorporeal cardiopulmonary circulation. The Ig can also be used in

XX diagnosing the presence and location of a thrombus, or certain types of

XX cancer cells which develop GPIIb/IIIa on their surfaces, for the

XX detection of GPIIb/IIIa antigens or for isolating platelets.

XX Sequence 107 AA;

```
RESULT 11
AAW58515
ID AAW58515 standard; Protein: 107 AA.
XX
AC AAW58515;
XX
DT 18-AUG-1998 (first entry)
XX
DE Protein SEQ ID NO:87 from US5770196.
XX
KW Humanised; human; mouse; CD5; anti-CD5 antibody; immunoglobulin;
KW depletion; cytotoxic; immunconjugate; fusion protein; psoriasis;
KW autoimmune disease; rheumatoid arthritis; type I diabetes.
XX
OS Unidentified.
XX
PN US5770196-A.
XX
PD 23-JUN-1998.
XX
PF 07-JUN-1995; 95US-0472788.
XX
PR 23-JUN-1993; 93US-0082842.
PR 13-DEC-1991; 91US-0808464.
PR 14-DEC-1992; 92WO-US10906.
PR 07-JUN-1995; 95US-0472788.
XX
PA (XOMA ) XOMA CORP.
XX
PI Studnicka GM;
XX
DR WPI: 1998-376744/32.
XX
PT Depletion of CD5-positive cells in vivo - using anti-CD5 antibodies
PT with humanised variable regions
XX
PS Disclosure: Column 93-94; 77pp: English.
XX
CC A method has been developed of depleting CD5+ cells in an animal. The
CC method comprises administering a cytotoxic protein containing a modified
CC immunoglobulin (Ig) variable domain, where the protein is an anti-CD5 Ig
CC molecule or an immunconjugate or fusion protein containing an anti-CD5
CC Ig molecule, and where the modified Ig variable domain comprises at
CC least one of (a) a modified light chain variable region (see AAW58478 or
CC AAW58480), and (b) a modified heavy chain variable region (see AAW58479
CC or AAW58481), where AAW58478 and AAW58479 are humanised forms of the H65
CC light and heavy chain variable domains with low risk amino acid
CC substitutions [i.e. low risk of reducing antigen-binding specificity.]
CC and AAW58480 and AAW58481 are humanised forms of the H65 light and heavy
CC chain variable domains with moderate risk amino acid substitutions and
CC are present in humanised H65 antibody he3 (ATCC HB 11206). The method is
CC useful for treating autoimmune diseases, especially systemic lupus
CC erythematosus, rheumatoid arthritis, psoriasis or type I diabetes. The
CC present sequence is given in the sequence listing of the present
CC invention but is not mentioned further within the specification.
XX
SQ Sequence 107 AA;
```

```
Query Match 100.0%; Score 61; DB 19; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 PWTGGGTRL 10
DB 95 PWTGGGTRL 104
```

```
RESULT 12
AAW58505
ID AAW58505 standard; Protein: 107 AA.
```

```
XX
AC AAW58505;
XX
DT 18-AUG-1998 (first entry)
XX
DE H65 protein sequence SEQ ID NO:26 from Fig 6.
XX
KW Humanised; human; mouse; CD5; anti-CD5 antibody; immunoglobulin;
KW depletion; cytotoxic; immunconjugate; fusion protein; psoriasis;
KW autoimmune disease; rheumatoid arthritis; type I diabetes.
XX
OS Mus sp.
XX
PN US5770196-A.
XX
PD 23-JUN-1998.
XX
PF 07-JUN-1995; 95US-0472788.
XX
PR 23-JUN-1993; 93US-0082842.
PR 13-DEC-1991; 91US-0808464.
PR 14-DEC-1992; 92WO-US10906.
PR 07-JUN-1995; 95US-0472788.
XX
PA (XOMA ) XOMA CORP.
XX
PI Studnicka GM;
XX
DR WPI: 1998-376744/32.
XX
PT Depletion of CD5-positive cells in vivo - using anti-CD5 antibodies
PT with humanised variable regions
XX
PS Example 6; Column 61-64; 77pp: English.
XX
CC A method has been developed of depleting CD5+ cells in an animal. The
CC method comprises administering a cytotoxic protein containing a modified
CC immunoglobulin (Ig) variable domain, where the protein is an anti-CD5 Ig
CC molecule or an immunconjugate or fusion protein containing an anti-CD5
CC Ig molecule, and where the modified Ig variable domain comprises at
CC least one of (a) a modified light chain variable region (see AAW58478 or
CC AAW58480), and (b) a modified heavy chain variable region (see AAW58479
CC or AAW58481), where AAW58478 and AAW58479 are humanised forms of the H65
CC light and heavy chain variable domains with low risk amino acid
CC substitutions [i.e. low risk of reducing antigen-binding specificity.]
CC and AAW58480 and AAW58481 are humanised forms of the H65 light and heavy
CC chain variable domains with moderate risk amino acid substitutions and
CC are present in humanised H65 antibody he3 (ATCC HB 11206). The method is
CC useful for treating autoimmune diseases, especially systemic lupus
CC erythematosus, rheumatoid arthritis, psoriasis or type I diabetes. The
CC present sequence represents an H65 sequence from the present invention.
XX
SQ Sequence 107 AA;
```

```
Query Match 100.0%; Score 61; DB 19; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 PWTGGGTRL 10
DB 95 PWTGGGTRL 104
```

```
RESULT 13
AAW58478
ID AAW58478 standard; Protein: 107 AA.
XX
AC AAW58478;
XX
DT 18-AUG-1998 (first entry)
```

```
DE Humanised anti-CD5 immunoglobulin H65 light chain variable region #1.
```

KW		Humanised:	human;	mouse;	CD5; anti-CD5 antibody; immunoglobulin;	
KV		depletion;	Cytotoxic;	immunconjugate;	fusion protein; psoriasis;	
KM		autoimmune disease;	rheumatoid arthritis;	Type I diabetes.		
XX		Synthetic.				
OS	Mus sp.	Homo sapiens.				
PN	US5770196-A.					
PD	23-JUN-1998.					
PF	07-JUN-1995;	95U5-0472788.				
PR	23-JUN-1993;	93U5-0082842.				
PR	13-DEC-1991;	91U5-0808464.				
PR	14-DEC-1992;	92WO-USI0906.				
PS	07-JUN-1995;	95U5-0472788.				
PA	(XOMA) XOMA CORP.					
PI	Studnicka GM;					
DR	WI: 1998-376744/32.					
PT	Depletion of CD5-positive cells in vivo - using anti-CD5 antibodies with humanised variable regions					
PS	Claim 1; Column 63-64; 77pp; English.					
CC	A method has been developed of depleting CD5+ cells in an animal. The method comprises administering a cytotoxic protein containing a modified immunoglobulin (Ig) variable domain, where the protein is an anti-CD5 Ig molecule or an immunoconjugate or fusion protein containing an anti-CD5 Ig molecule, and where the modified Ig variable domain comprises at least one of (a) a modified light chain variable region (see AAW58478 or AAW58480), and (b) a modified heavy chain variable region (see AAW58479 or AAW58491), where AAW58478 and AAW58479 are humanised forms of the H65 light and heavy chain variable domains with low risk amino acid substitutions [i.e., low risk of reducing antigen-binding specificity.] and AAW58480 and AAW58481 are humanised forms of the H65 light and heavy chain variable domains with moderate risk amino acid substitutions and are present in humanised H65 antibody he3 (ATCC HB 11205). The method is useful for treating autoimmune diseases, especially systemic lupus erythematosus, rheumatoid arthritis, psoriasis or type I diabetes. The present sequence represents a specifically claimed humanised anti-CD5 immunoglobulin H65 light chain variable region. Sequence 107 AA; SQ					
OY	Query Match	100.0%;	Score 61;	DB 19;	Length 107;	
Dd	Best Local Similarity	100.0%;	Pred. NO. 0.035;			
	Matches 10; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0.	
QY	1 PWTGEGGTKL 10 					
	Db 95 pwtfggtkl 104					
RESULT 14						
ID	AAAR21292					
AC	.AAR21292;					
DT	21-MAY-1992 (first entry)					
DE	Murine VL kappa group VI chain "g", specific for phox.					
Fd,	bacteriophage; gene III; filamentous; phagemid; capsid; coat;					
pilus; gdp;	binding; adsorption; gene VIII; diverse repertoire;					

```

KW specific binding pairs; replicable genetic display package.
XX Synthetic.
OS
XX
FH Key Location/Qualifiers
FT Binding-site 24..33 /label= CDR1
FT Binding-site 49..55 /label= CDR2
FT Binding-site 88..96 /label= CDR3
FT FT /note=" D-X-G-X-X motif "
XX
XX W09201047-A.
PN
XX
PD 23-JAN-1992.
XX
PE 10-JUL-1991; 91WO-GB01134.
XX
PR 15-MAY-1991; 91GB-0010549.
PR 10-JUL-1990; 90GB-0015198.
PR 19-OCT-1990; 90GB-0022845.
PR 12-NOV-1990; 90GB-0024503.
PR 06-MAR-1991; 91GB-0004744.
XX
PA (CAMP-) CAMBRIDGE ANTIBODY.
PA (MEDI-) MED RES COUNCIL.
XX
PI McCafferty J, Pope AR, Johnson KS, Hoogenboom HRJ, Griffiths AD;
PI Jackson RH, Holliger KP, Marks JD, Clackson TP, Chiswell DJ;
PI Winter GP, Bonnett TP;
XX
XX MPI; 1992-056862/07.
XX
PT Producing members of specific binding pairs - by expression in
PT recombinant host cells with a secreting replicable genetic
PT display package.
XX
XX Example 21; Fig 24; 209pp; English.
XX
CC The VK sequence is one of seven (AAR21286-92) found to be expressed
CC from a single chain Fv library from an immunised mouse. The libra-
CC ry produces a diverse repertoire of antibody fragments specific for
CC 2-phenyl-5-oxazalone (phox). It was prep'd using cDNA generated
CC from mRNA from mice immunised with phox coupled to chicked serum
CC albumin. The VH and VL kappa sequences were separately amplified
CC by PCR (see AAQ24474-84) and ligated into ficator2 (see AAQ23463) for ex-
CC pression on the phage surface as fusions with gene III. The result-
CC ing library of clones was diverse. Twenty three hapten binding
CC clones were sequenced revealing eight different VH genes (A-H) (see
CC AAR1244-71) in a variety of pairings with the seven different VK
CC genes (a-g). Of the twenty three clones sequenced, only one was of
CC type "g", and was an "ox-like" gene. (see Berek et al., Nature 316
CC 412-418, 1985). It contains the DXGX motif in CDR3, the central
CC Gly of which is needed to create a cavity for phox. Most of the
CC clones were VK-d combinations. The Kd of VH-B/VK-b for phox-GABA
CC was 10 uM, one of the highest values found. This suggests that
CC phage beating scfv fragments having weak affinities can be selected
CC with antigen, probably due to the avidity of the multiple antibody
CC heads on the phage.
CC See also AAR21260-307, 309-311; AAR22450, 565-581.
XX
XX Sequence 108 AA:
SQ

```

```
RESULT 15
AAR21310
ID AAR21310 standard; Protein: 108 AA.
XX
AC AAR21310;
XX
DT 21-MAY-1992 (first entry)
XX
DE Light chain of M1f clone.
XX
KW Fd: bacteriophage; gene III; filamentous; phagemid; capsid; coat;
KW plus; g3p; binding; adsorption; gene VIII; diverse repertoire;
KW specific binding pairs; replicable genetic display package.
XX
OS Synthetic.
XX
FH Key
FH Region 25..34
FT /label- CDR1
FT Region 50..56
FT /label- CDR2
FT Region 89..96
FT /label- CDR3
XX
XX W09201047-A.
XX
XX 23-JAN-1992.
XX
XX 10-JUL-1991; 91WO-GB01134.
XX
XX 15-MAY-1991; 91GB-0010549.
XX
XX 10-JUL-1990; 90GB-0015198.
XX
XX 19-OCT-1990; 90GB-0022845.
XX
XX 12-NOV-1990; 90GB-0024503.
XX
XX 06-MAR-1991; 91GB-0004744.
XX
XX (CMB-) CAMBRIDGE ANTIBODY.
XX
XX (MEDI-) MED RES COUNCIL.
XX
XX McCafferty J, Pope AR, Johnson KS, Hoogenboom HRJ, Griffiths AD;
XX Jackson RH, Holliger KP, Marks JD, Jackson TP, Chiswell DJ;
XX Winter GP, Bonnett TP;
XX
XX WPI: 1992-056862/07.
XX
XX Producing members of specific binding pairs - by expression in
XX recombinant host cells with a secreting replicable genetic
XX display package.
XX
XX Example 46; Fig 52; 109pp; English.
XX
XX The sequence is the light chain of clone M1f encoding an scFv frag-
XX ment specific for both hen and turkey egg lysozyme (HEL and TEL).
XX The DNA encoding the chain was amplified from a cDNA library prepd.
XX from the spleen of an unimmunised mouse. The corresponding heavy
XX chain was also amplified from an existing construct, PSW1-VHD1.3
XX (Ward et al, 1989). The two fragments were assembled via a linker
XX to prepare an scFv construct which was ligated into the fdCAT2
XX vector for expression on the surface of fd bacteriophage. In this
XX way, the VL domain was replaced by a library of VL domains to allow
XX for selection of a broader range of antibody specificities. Several
XX clones were isolated which bound to TEL (the parent antibody D1.3
XX binds exclusively to HEL). The sequences of the light chains of
XX two of these clones, M1 and M21 are given in AAR21310 and AAR21311
XX respectively. The D1.3 light chain is given in AAR21309.
XX See also AAR21260-307, 308-312, AAR22450, AAR22565, AAR22567-81.
XX
XX SQ Sequence 108 AA;
```

```
Query Match 100.0%; Score 61; DB 13; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.035;
```

```
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PWTFGGTRL 10
| | | | |
Db 95 pwtfggtrkl 104
```

```
Search completed: August 20, 2002, 13:15:11
Job time: 105 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:16:03 ; Search time 46.28 Seconds
(without alignments)
20.763 Million cell updates/sec

Title: US-09-824-286-4_COPY_95_104
Perfect score: 61
Sequence: 1 PMTFGGGCTKL 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	77	2 A28840	Ig kappa chain V r
2	61	100.0	79	2 A28840	Ig kappa chain V r
3	61	100.0	88	2 PL0261	Ig kappa chain V r
4	61	100.0	95	2 S26333	Ig kappa chain V r
5	61	100.0	96	2 G38601	Ig kappa chain V r
6	61	100.0	98	2 S19974	Ig kappa chain V r
7	61	100.0	103	2 S18731	Ig kappa chain V r
8	61	100.0	106	2 PL0259	Ig kappa chain V r
9	61	100.0	106	2 PL0259	Ig kappa chain V r
10	61	100.0	106	2 PL0262	Ig kappa chain V r
11	61	100.0	107	2 S09964	Ig kappa chain V r
12	61	100.0	107	2 PL0268	Ig kappa chain V r
13	61	100.0	107	2 S09968	Ig kappa chain V r
14	61	100.0	108	2 S19112	Ig kappa chain V r
15	61	100.0	108	2 S38862	Ig kappa chain V r
16	61	100.0	108	2 S69900	Ig kappa chain V r
17	61	100.0	109	2 PH0093	Ig kappa chain V r
18	61	100.0	109	2 S26336	Ig kappa chain V r
19	61	100.0	110	2 PH0090	Ig kappa chain V r
20	61	100.0	111	1 KVM880	Ig kappa chain V r
21	61	100.0	111	2 E38740	Ig kappa chain V r
22	61	100.0	111	2 A38740	Ig kappa chain V r
23	61	100.0	111	2 G38740	Ig kappa chain V r
24	61	100.0	111	2 C38740	Ig kappa chain V r
25	61	100.0	112	2 A27887	Ig kappa chain V r
26	61	100.0	112	2 C27887	Ig kappa chain V r
27	61	100.0	112	2 A49715	Ig kappa chain V r
28	61	100.0	112	2 B31485	Ig kappa chain V r
29	61	100.0	112	2 A36259	Ig kappa chain V r

30	61	100.0	112	2 PL0273	Ig kappa chain V r
31	61	100.0	112	2 S38716	Ig kappa chain V r
32	61	100.0	113	2 E30560	Ig kappa chain V r
33	61	100.0	114	2 PT0358	Ig kappa chain V r
34	61	100.0	118	2 PT0359	Ig kappa chain V r
35	61	100.0	121	2 S67944	Ig kappa chain V r
36	61	100.0	122	2 A29380	Ig kappa chain V r
37	61	100.0	126	2 A34904	Ig kappa chain V r
38	61	100.0	127	2 S04577	Ig kappa chain V r
39	61	100.0	127	2 S04577	Ig kappa chain V r
40	61	100.0	128	2 A47159	Ig kappa chain V r
41	61	100.0	130	1 KVM5M4	Ig kappa chain V r
42	61	100.0	131	1 KVM5M6	Ig kappa chain V r
43	61	100.0	131	2 B34904	Ig kappa chain V r
44	61	100.0	131	2 C34904	Ig kappa chain V r
45	61	100.0	131	2 S52449	Ig kappa chain V r

ALIGNMENTS

RESULT 1
G28840
Ig kappa chain V region (HP25) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 23-Jul-1999
C:Accession: G28840; 025114
R:Ollier, P.; Rocca-Serra, J.; Somme, G.; Theze, J.; Fougereau, M.
EMBO J. 4, 3681-3688, 1985
A:Title: The idiotypic network and the internal image: possible regulation of a germ-
A:Reference number: A91028; MUID:86136012
A:Accession: G28840
A:Molecule type: mRNA
A:Residues: 1-77 <OH>
A:Cross-references: GB:X03389; NID:952174; PIDN:CAA27120.1; PID:952175
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: heterotrimer; Immunoglobulin

Query Match 100.0%; Score 61; DB 2; Length 77;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PMTFGGGCTKL 10
DB 65 PMTFGGGCTKL 74

RESULT 2
A28840
Ig kappa chain V region (HP21) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change 23-Jul-1999
C:Accession: A28840; M25114
R:Ollier, P.; Rocca-Serra, J.; Somme, G.; Theze, J.; Fougereau, M.
EMBO J. 4, 3681-3688, 1985
A:Title: The idiotypic network and the internal image: possible regulation of a germ-
A:Reference number: A91028; MUID:86136012
A:Accession: A28840
A:Molecule type: mRNA
A:Residues: 1-79 <OH>
A:Cross-references: GB:X03387; NID:952168; PIDN:CAA27118.1; PID:952169
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: heterotrimer; Immunoglobulin

Query Match 100.0%; Score 61; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PMTFGGGCTKL 10
DB 67 PMTFGGGCTKL 76

RESULT 3
PL0261
Ig kappa chain V region (anti-DNA, DP15VK) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-May-1997
C:Accession: PL0261
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein, A.
J. Exp. Med. 171, 265-297, 1990
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic
A:Reference number: PL0231; MUID:90111618
A:Accession: PL0261
A:Molecule type: mRNA
A:Residues: 1-88 <SHL>
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:1-5/Region: framework 1
F:6-16/Region: complementarity-determining 1
F:17-31/Region: framework 2
F:32-38/Region: complementarity-determining 2
F:39-70/Region: framework 3
F:71-79/Region: complementarity-determining 3
F:80-88/Region: framework 4

Query Match 100.0%; Score 61; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 0.0032;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGCGGCKL 10
|||||
DB 77 PWTGCGGCKL 86

RESULT 4
S26333
Ig kappa chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Jun-2000
C:Accession: S26333
R:Stark, S.E.; Caton, A.J.
J. Exp. Med. 174, 613-624, 1991
A:Title: Antibodies that are specific for a single amino acid interchange in a protein
A:Reference number: S26309; MUID:91341421
A:Accession: S26333
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-95 <STRA>
A:Cross-references: EMBL:X59181; NID:952312; PID:91334061
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:6-80/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGCGGCKL 10
|||||
DB 85 PWTGCGGCKL 94

RESULT 5
G38601
Ig kappa chain V region (4A9) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 30-Aug-1991 #sequence_revision 30-Aug-1991 #text_change 23-Jul-1999
C:Accession: G38601
R:Goshorn, S.C.; Retzel, E.; Jemerson, R.
J. Biol. Chem. 266, 2134-2142, 1991
A:Title: Common structural features among monoclonal antibodies binding the same antigen

A:Reference number: A38601; MUID:91115823
A:Accession: G38601
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-96 <GOS>
A:Cross-references: GB:M57984; NID:9196414; PID:AAA63365.1; PID:9196415
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 61; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGCGGCKL 10
|||||
DB 83 PWTGCGGCKL 92

RESULT 6
S19974
Ig kappa chain V region (M-T406) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 21-Jan-2000
C:Accession: S19974
R:Weissenhorn, W.; Rietmueller, G.; Weiss, E.M.; Rieber, E.P.
submitted to the EMBL Data Library, March 1992
A:Description: Structural characterization of CD4 mab.
A:Reference number: S19963
A:Accession: S19974
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-98 <WEI>
A:Cross-references: EMBL:X65096; NID:952294; PID:CAA6224.1; PID:952295
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:1-80/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGCGGCKL 10
|||||
DB 85 PWTGCGGCKL 94

RESULT 7
S18731
Ig kappa chain V-J region (MSI-N17) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 05-Dec-1998 #text_change 23-Jul-1999
C:Accession: S18731
R:Hirama, T.; Takeshita, S.; Yoshida, Y.; Yamaagishi, H.
Immunol. Lett. 27, 19-24, 1991
A:Title: Structure of extrachromosomal circular DNAs generated by immunoglobulin high
A:Reference number: S18731; MUID:91209891
A:Accession: S18731
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-103 <HIR>
A:Cross-references: EMBL:X54753; NID:955294; PID:CAA38555.1; PID:955295
A:Experimental source: spleen, strain BALB/c-nu/nu
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1990
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (chain disulfide bonds. In some cases, such as Iga and Igm, the subunits associate into
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:11-85/Domain: immunoglobulin homology <IMM>
F:18-83/Disulfide bonds: #status predicted

Query Match 100.0%; Score 61; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.0037;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10
|||||
DB 90 PWTGGGCTKL 99

RESULT 8

PL0260
Ig kappa chain V region (anti-DNA, DP7VK) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 21-Jan-2000
C:Accession: PL0260
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein, A.
J. Exp. Med. 171, 265-297, 1990
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic
A:Reference number: PL0231; MVID:90111618
A:Accession: PL0260
A:Molecule type: mRNA
A:Residues: 1-106 <SHL>
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:1-23/Region: framework 1
F:16-90/Domain: immunoglobulin homology <IMM>
F:24-34/Region: complementarity-determining 1
F:35-49/Region: framework 2
F:50-56/Region: complementarity-determining 2
F:57-88/Region: framework 3
F:89-97/Region: complementarity-determining 3
F:98-106/Region: framework 4

Query Match 100.0%; Score 61; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10
|||||
DB 95 PWTGGGCTKL 104

RESULT 9

PL0259
Ig kappa chain V region (anti-DNA, DP11VK) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 21-Jan-2000
C:Accession: PL0259
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein, A.
J. Exp. Med. 171, 265-297, 1990
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic
A:Reference number: PL0231; MVID:90111618
A:Accession: PL0259
A:Molecule type: mRNA
A:Residues: 1-106 <SHL>
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:1-23/Region: framework 1
F:16-90/Domain: immunoglobulin homology <IMM>
F:24-34/Region: complementarity-determining 1
F:35-49/Region: framework 2
F:50-56/Region: complementarity-determining 2
F:57-88/Region: framework 3
F:89-97/Region: complementarity-determining 3
F:98-106/Region: framework 4

Query Match 100.0%; Score 61; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10

|||||
DB 95 PWTGGGCTKL 104

RESULT 10

PL0262
Ig kappa chain V region (anti-DNA, 60VK) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 21-Jan-2000
C:Accession: PL0262
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein, A.
J. Exp. Med. 171, 265-297, 1990
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic
A:Reference number: PL0231; MVID:90111618
A:Accession: PL0262
A:Molecule type: mRNA
A:Residues: 1-106 <SHL>
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:1-23/Region: framework 1
F:16-90/Domain: immunoglobulin homology <IMM>
F:24-34/Region: complementarity-determining 1
F:35-49/Region: framework 2
F:50-56/Region: complementarity-determining 2
F:57-88/Region: framework 3
F:89-97/Region: complementarity-determining 3
F:98-106/Region: framework 4

Query Match 100.0%; Score 61; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10
|||||
DB 95 PWTGGGCTKL 104

RESULT 11

S09964
Ig kappa chain V-J region (105-2H) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 21-Jan-2000
C:Accession: S09964
R:Reininger, L.; Shibata, T.; Ozaki, S.; Shirai, T.; Jaton, J.C.; Izui, S.
Eur. J. Immunol. 20, 771-777, 1990
A:Title: Variable region sequences of pathogenic anti-mouse red blood cell autoantibodies
A:Reference number: S09955; MVID:90269328
A:Accession: S09964
A:Molecule type: mRNA
A:Residues: 1-107 <REI>
A:Cross-references: EMBL:X51852; NID:g55393; PIDN:CA36145.1; PID:g930229
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10
|||||
DB 95 PWTGGGCTKL 104

RESULT 12

PL0268
Ig kappa chain V region (anti-DNA, DP9VK and DP17VK) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 21-Jan-2000
C:Accession: PL0268
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein, A.

J. Exp. Med. 171, 265-297, 1990
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic
A:Reference number: PL0231; MUID:90111618
A:Accession: PL0268
A:Molecule type: mRNA
A:Residues: 1-107 <SHL>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:1-23/Region: framework 1
F:16-90/Domain: immunoglobulin homology <IMM>
F:24-34/Region: complementarity-determining 1
F:35-49/Region: complementarity-determining 2
F:50-56/Region: complementarity-determining 3
F:57-68/Region: framework 3
F:89-97/Region: complementarity-determining 3
F:98-107/Region: framework 4

Query Match 100.0%; Score 61; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCKL 10
|||
Db 95 PWTGGGCKL 104

RESULT 13
S09968
Ig kappa chain V-J region (34-2B) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 21-Jan-2000
R:Accession: S09968
R:Reininger, L.; Shihata, T.; Ozaki, S.; Shirai, T.; Jaton, J.C.; Izui, S.
Eur. J. Immunol. 20, 771-777, 1990
A:Title: Variable region sequences of pathogenic anti-mouse red blood cell autoantibodies
A:Reference number: S09955; MUID:90263328
A:Accession: S09968
A:Molecule type: mRNA
A:Residues: 1-107 <REI>
A:Cross-references: EMBL:X51856; NID:955403; PIDN:CA36149.1; PID:9930233
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCKL 10
|||
Db 95 PWTGGGCKL 104

RESULT 14
S19112
Ig kappa chain V region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jan-2000
R:Accession: S19112
R:Mylvaganam, S.E.; Paterson, Y.; Kaiser, K.; Bowdish, K.; Getzoff, E.D.
J. Mol. Biol. 221, 455-462, 1991
A:Title: Biochemical implications from the variable gene sequences of an anti-cytochrome
forms.
A:Reference number: S17586; MUID:92015240
A:Accession: S19112
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-108 <MTL>
A:Cross-references: EMBL:X60684
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin

F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCKL 10
|||
Db 95 PWTGGGCKL 104

RESULT 15
S38862
Ig kappa chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 21-Jan-2000
R:Accession: S38862
R:Fischer, R.; Voss, A.; Hunziker, W.; Stierhof, Y.D.; Kreuzaler, F.
submitted to the EMBL Data Library, August 1993
A:Description: Production and cloning of TMV-specific monoclonal antibodies.
A:Reference number: S37200
A:Accession: S38862
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-108 <FTS>
A:Cross-references: EMBL:X75854; NID:9429109; PID:9429110
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCKL 10
|||
Db 96 PWTGGGCKL 105

Search completed: August 20, 2002, 13:16:04
Job time: 158 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:52 ; Search time 21.48 seconds

(without alignments)
18.026 Million cell updates/sec

Title: US-09-824-286-4_COPY_95_104
Perfect score: 61
Sequence: 1 PWTGCGTKL 10

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	100.0	111	1	KV3A_MOUSE
2	61	100.0	111	1	KV3C_MOUSE
3	61	100.0	130	1	KV5G_MOUSE
4	61	100.0	131	1	KV3I_MOUSE
5	55	90.2	110	1	KV3P_MOUSE
6	55	90.2	111	1	KV3E_MOUSE
7	55	90.2	111	1	KV3O_MOUSE
8	55	90.2	111	1	KV3Q_MOUSE
9	55	90.2	112	1	KV3G_MOUSE
10	55	90.2	132	1	KV3F_MOUSE
11	54	88.5	107	1	KV6A_MOUSE
12	52	85.2	108	1	KV5Q_MOUSE
13	52	85.2	108	1	KV5R_MOUSE
14	52	85.2	108	1	KV5S_MOUSE
15	52	85.2	108	1	KV5T_MOUSE
16	52	85.2	108	1	KV5U_MOUSE
17	52	85.2	111	1	KV3D_MOUSE
18	52	85.2	111	1	KV3H_MOUSE
19	52	85.2	111	1	KV3J_MOUSE
20	52	85.2	112	1	KV3L_MOUSE
21	52	85.2	112	1	KV2D_MOUSE
22	52	85.2	112	1	KV3B_MOUSE
23	52	85.2	113	1	KV2E_MOUSE
24	52	85.2	113	1	KV2F_MOUSE
25	52	85.2	113	1	KV2G_MOUSE
26	50	82.0	108	1	KV1B_HUMAN
27	50	82.0	108	1	KV1R_HUMAN
28	50	82.0	129	1	KV3M_HUMAN
29	50	82.0	134	1	KV4C_HUMAN
30	49	80.3	110	1	KV1C_MOUSE
31	49	80.3	111	1	KV6C_HUMAN
32	49	80.3	112	1	KV6B_HUMAN
33	49	80.3	129	1	KV1B_MOUSE

ALIGNMENTS

34	49	80.3	129	1	LV1D_MOUSE	P01726 mus musculu
35	49	80.3	129	1	LV1E_MOUSE	P01727 mus musculu
36	49	80.3	130	1	LV1G_HUMAN	P06316 homo sapien
37	47	77.0	108	1	KV5J_MOUSE	P01643 mus musculu
38	47	77.0	108	1	KV5K_MOUSE	P01644 mus musculu
39	47	77.0	108	1	KV5L_MOUSE	P01645 mus musculu
40	47	77.0	108	1	KV5M_MOUSE	P01646 mus musculu
41	47	77.0	108	1	KV5N_MOUSE	P01647 mus musculu
42	47	77.0	108	1	KV5O_MOUSE	P01648 mus musculu
43	47	77.0	108	1	KV5P_MOUSE	P01649 mus musculu
44	47	77.0	111	1	KV3R_MOUSE	P01670 mus musculu
45	47	77.0	129	1	KV3H_HUMAN	P04207 homo sapien

RESULT 1						
ID	KV3A_MOUSE	STANDARD;	PRT;	111 AA.		
AC	P01654:					
DT	21-JUL-1986 (Rel. 01, Created)					
DT	21-JUL-1986 (Rel. 01, Last sequence update)					
DT	15-JUL-1999 (Rel. 38, Last annotation update)					
DE	Ig kappa chain V-III region PC 2880/PC 1229.					
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.					
OX	NCBI_Taxid=10090;					
RN	[1]					
RP	SEQUENCE.					
RX	MEDLINE=79073152; PubMed=103003;					
RA	Weigert M., Gattatran L., Loh E., Schilling J., Hood L.E.;					
RT	"Rearrangement of genetic information may produce immunoglobulin					
RT	diversity."					
RL	Nature 276:785-790(1978).					
CC	1- MISCELLANEOUS: THE PC 2880 AND PC 1229 SEQUENCES ARE IDENTICAL.					
CC	PIR: A01300; KWS80.					
DR	HSSP: P01789; IMCP.					
DR	Interpro: IPR003006; Ig_MHC.					
DR	Interpro: IPR003596; Ig_V.					
DR	PIR: P00047; Ig; 1.					
DR	SMART: SM00406; IgV; 1.					
KW	Immunoglobulin V region.					
FT	DOMAIN 1					FRAMEWORK-1.
FT	DOMAIN 23					COMPLEMENTARITY-DETERMINING-1.
FT	DOMAIN 39					FRAMEWORK-2.
FT	DOMAIN 54					COMPLEMENTARITY-DETERMINING-2.
FT	DOMAIN 61					FRAMEWORK-3.
FT	DOMAIN 93					COMPLEMENTARITY-DETERMINING-3.
FT	DOMAIN 102					FRAMEWORK-4.
FT	DISULFID 23					BY SIMILARITY.
FT	NON_TER 111					
SO	SEQUENCE 111 AA; 11980 MW; AFAC6A9D26FC12D CRC64;					
Query Match		100.0%;	Score 61;	DB 1;	Length 111;	
Best Local Similarity		100.0%;	Pred. No. 0.0011;			
Matches 10;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
OY	1 PWTGCGTKL 10					
DB	99 PWTGCGTKL 108					
RESULT 2						
ID	KV3C_MOUSE	STANDARD;	PRT;	111 AA.		
AC	P01656:					
DT	21-JUL-1986 (Rel. 01, Created)					
DT	21-JUL-1986 (Rel. 01, Last sequence update)					
DT	15-JUL-1999 (Rel. 38, Last annotation update)					
DE	Ig kappa chain V-III region MOPC 70.					

OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 FN [1]
 RP MEDLINE=67056897; PubMed=4162931;
 RA Gray W.R., Dreyer W.J., Hood L.E.;
 RT "Mechanism of antibody synthesis: size differences between mouse
 kappa chains.";
 RL Science 155:465-467(1967).
 CC -1 MISCELLANEOUS: THIS IS A BENCE-JONES PROTEIN.
 DR PIR: A01930; KWSM6;
 DR HSSP: P01789; IMCP;
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig; 1.
 DR SMART: SM00406; IgV; 1.
 KW Immunoglobulin V region; Bence-Jones protein.
 FT DOMAIN 1 23
 FT DOMAIN 24 38
 FT DOMAIN 39 53
 FT DOMAIN 54 60
 FT DOMAIN 61 92
 FT DOMAIN 93 101
 FT DOMAIN 102 111
 FT DISULFID 23-92
 FT NON_TER 111
 FT SEQUENCE 111 AA; 11904 MW; 4FE7ABC9DF0FC125 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 61; DB 1; Length 111;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PWTFGGCTKL 10
 ID 99 PWTFGGCTKL 108

RESULT 3
 ID KVS_MOUSE STANDARD; PRT; 130 AA.
 AC P01639; P01640;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ig kappa chain V-V region MOPC 41 precursor.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 FN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=79221900; PubMed=111146;
 RA Seidman J.G., Max F.E., Leder P.;
 RT "A kappa-immunoglobulin gene is formed by site-specific recombination
 without further somatic mutation.";
 RL Nature 280:370-375(1979).
 FN [2]
 RP SEQUENCE OF 1-33.
 RA MEDLINE=77148916; PubMed=403522;
 RA Burstein Y., Schechter I.;
 RT "Amino acid sequence of the NH2-terminal extra piece segments of the
 RT precursors of mouse immunoglobulin lambda1-type and kappa-type light
 RT chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 74:716-720(1977).
 FN [3]
 RP SEQUENCE OF 23-130.
 RA MEDLINE=67056897; PubMed=4162931;
 RA Gray W.R., Dreyer W.J., Hood L.;
 RT "Mechanism of antibody synthesis: size differences between mouse
 RT kappa chains.";

RL Science 155:465-467(1967).
 CC -1 MISCELLANEOUS: THIS PRECURSOR WAS SYNTHESIZED IN A CELL-FREE
 CC SYSTEM DIRECTED BY mRNA ISOLATED FROM MYELOMA POLYSOMES.
 CC -1 MISCELLANEOUS: THIS IS A BENCE-JONES PROTEIN.
 DR PIR: A01922; KWSM4.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig; 1.
 DR SMART: SM00406; IgV; 1.
 KW Immunoglobulin V region; Signal; Bence-Jones protein.
 FT SIGNAL 1 22
 FT CHAIN 23 130
 FT DOMAIN 23 45
 FT DOMAIN 46 56
 FT DOMAIN 57 71
 FT DOMAIN 72 78
 FT DOMAIN 79 110
 FT DOMAIN 111 119
 FT DOMAIN 120 129
 FT DISULFID 45-110
 FT VARIANT 1 2
 FT NON_TER 130
 FT SEQUENCE 130 AA; 14311 MW; 5EFE0FE71D5F1BEC CRC64;

Query Match
 Best Local Similarity 100.0%; Score 61; DB 1; Length 130;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PWTFGGCTKL 10
 DB 117 PWTFGGCTKL 126

RESULT 4
 ID KVS_MOUSE STANDARD; PRT; 131 AA.
 AC P01661;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ig kappa chain V-III region MOPC 63 precursor.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 FN [1]
 RP SEQUENCE OF 1-35.
 RA MEDLINE=78235887; PubMed=98179;
 RA Burstein Y., Schechter I.;
 RT "Primary structures of N-terminal extra peptide segments linked to
 RT the variable and constant regions of immunoglobulin light chain
 RT precursors: implications on the organization and controlled
 RT expression of immunoglobulin genes.";
 RL Biochemistry 17:2392-2400(1978).
 FN [2]
 RP SEQUENCE OF 21-131.
 RA MEDLINE=73140225; PubMed=4691517;
 RA McKean D.J., Potter M., Hood L.E.;
 RT "Mouse immunoglobulin chains. Pattern of sequence variation among
 RT kappa chains with limited sequence differences.";
 RL Biochemistry 12:760-771(1973).
 FN [3]
 RP REVISIONS.
 RA MEDLINE=79012520; PubMed=99744;
 RA McKean D.J., Bell M., Potter M.;
 RT "Mechanisms of antibody diversity: multiple genes encode structurally
 RT related mouse kappa variable regions.";
 RL Proc. Natl. Acad. Sci. U.S.A. 75:3913-3917(1978).
 DR PIR: A01935; KWSM6.
 DR HSSP: P01789; IMCP.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_V.

DR Pfam: PF00047; 1g; 1.
 DR SMART: SM00406; IGV; 1.
 KW Immunoglobulin V region; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 131 IG KAPPA CHAIN V-III REGION MOPC 63.
 FT DOMAIN 21 43 FRAMEWORK-1.
 FT DOMAIN 44 58 COMPLEMENTARITY-DETERMINING-1.
 FT DOMAIN 59 73 FRAMEWORK-2.
 FT DOMAIN 74 80 COMPLEMENTARITY-DETERMINING-2.
 FT DOMAIN 81 112 FRAMEWORK-3.
 FT DOMAIN 113 121 COMPLEMENTARITY-DETERMINING-3.
 FT DOMAIN 122 131 FRAMEWORK-4.
 FT DISULFID 43 112 BY SIMILARITY.
 FT NON_TER 131 131
 SQ SEQUENCE 131 AA; 14291 MW; D212EC9F08DC880A CRC64;

Query Match 100.0%; Score 61; DB 1; Length 131;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTFCGGTKL 10
 DB 119 PWTFCGGTKL 128

RESULT 5
 KV3P_MOUSE STANDARD: PRT: 110 AA.
 AC P01667;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ig kappa chain V-III region PC 7210.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=79073152; PubMed=103003;
 RA Weigert M., Gattalitan L., Loh E., Schilling J., Hood L.E.;
 RT "Rearrangement of genetic information may produce immunoglobulin
 RT diversity."
 RL Nature 276:785-790(1978).
 RL PIR: D01937; KWS10.
 DR HSSP; P01789; IMCP.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003596; IG_V.
 DR Pfam: PF00047; 1g; 1.
 DR SMART: SM00406; IGV; 1.
 KW Immunoglobulin V region.
 FT DOMAIN 1 23 FRAMEWORK-1.
 FT DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.
 FT DOMAIN 39 53 FRAMEWORK-2.
 FT DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.
 FT DOMAIN 61 92 FRAMEWORK-3.
 FT DOMAIN 93 100 COMPLEMENTARITY-DETERMINING-3.
 FT DOMAIN 101 110 FRAMEWORK-4.
 FT DISULFID 23 92 BY SIMILARITY.
 FT NON_TER 110 110
 SQ SEQUENCE 110 AA; 11950 MW; 69F1A5CE886B1249 CRC64;

Query Match 90.2%; Score 55; DB 1; Length 110;
 Best Local Similarity 90.0%; Pred. No. 0.011;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PWTFCGGTKL 10
 DB 98 PWTFCGGTKL 107

RESULT 6
 KV3E_MOUSE STANDARD: PRT: 111 AA.
 AC P01657;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ig kappa chain V-III region PC 2413.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=79073152; PubMed=103003;
 RA Weigert M., Gattalitan L., Loh E., Schilling J., Hood L.E.;
 RT "Rearrangement of genetic information may produce immunoglobulin
 RT diversity."
 RL Nature 276:785-790(1978).
 RL PIR: A01932; KWS13.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003596; IG_V.
 DR Pfam: PF00047; 1g; 1.
 DR SMART: SM00406; IGV; 1.
 KW Immunoglobulin V region.
 FT DOMAIN 1 23 FRAMEWORK-1.
 FT DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.
 FT DOMAIN 39 53 FRAMEWORK-2.
 FT DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.
 FT DOMAIN 61 92 FRAMEWORK-3.
 FT DOMAIN 93 101 COMPLEMENTARITY-DETERMINING-3.
 FT DOMAIN 102 111 FRAMEWORK-4.
 FT DISULFID 23 92 BY SIMILARITY.
 FT NON_TER 111 111
 SQ SEQUENCE 111 AA; 11949 MW; ADB89EFC6EC746A9 CRC64;

Query Match 90.2%; Score 55; DB 1; Length 111;
 Best Local Similarity 90.0%; Pred. No. 0.011;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PWTFCGGTKL 10
 DB 99 PWTFCGGTKL 108

RESULT 7
 KV3O_MOUSE STANDARD: PRT: 111 AA.
 AC P01667;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ig kappa chain V-III region PC 6308.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=79073152; PubMed=103003;
 RA Weigert M., Gattalitan L., Loh E., Schilling J., Hood L.E.;
 RT "Rearrangement of genetic information may produce immunoglobulin
 RT diversity."
 RL Nature 276:785-790(1978).
 RL PIR: C01937; KWS08.
 DR HSSP; P01789; IMCP.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003596; IG_V.
 DR Pfam: PF00047; 1g; 1.
 DR SMART: SM00406; IGV; 1.
 KW Immunoglobulin V region.
 FT DOMAIN 1 23 FRAMEWORK-1.

```

FT DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 39 53 FRAMEWORK-2.
FT DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 61 92 FRAMEWORK-3.
FT DOMAIN 93 101 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 102 111 FRAMEWORK-4.
FT DISULFID 23 92 BY SIMILARITY.
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12071 MM; 7A4ADE4D6C256D29 CRC64;

Query Match
Best Local Similarity 90.2%; Score 55; DB 1; Length 111;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PWTFGSGTKL 10
DB 99 PWTFGSGTKL 108

RESULT 8
KV3Q_MOUSE STANDARD; PRT; 111 AA.
AC P01659;
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region PC 7769.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79073152; PubMed=103003;
RA Weigert M., Galtmaitan L., Loh E., Schilling J., Hood L.E.;
RT "Rearrangement of genetic information may produce immunoglobulin
RT diversity.";
RT Nature 276:785-790(1978).
RU PIR: E01937; KVM569.
DR HSSP; P01789; IMCP.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGV_1.
KM Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 39 53 FRAMEWORK-2.
FT DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 61 92 FRAMEWORK-3.
FT DOMAIN 93 101 FRAMEWORK-4.
FT DISULFID 23 92 BY SIMILARITY.
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12011 MM; 6FAA345279356829 CRC64;

Query Match
Best Local Similarity 90.2%; Score 55; DB 1; Length 111;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PWTFGSGTKL 10
DB 99 PWTFGSGTKL 108

RESULT 9
KV3G_MOUSE STANDARD; PRT; 112 AA.
AC P01659;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)

```

```

DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region TEPC 124.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=73140225; PubMed=4691517;
RA McKean D.J., Potter M., Hood L.E.;
RT "Mouse immunoglobulin chains. Pattern of sequence variation among
RT kappa chains with limited sequence differences.";
RT Biochemistry 12:760-771(1973).
DR PIR; A01933; KVM532.
DR HSSP; P01607; IREL.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGV_1.
KM Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 39 53 FRAMEWORK-2.
FT DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 61 92 FRAMEWORK-3.
FT DOMAIN 93 101 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 23 92 BY SIMILARITY.
FT NON_TER 112 112
SQ SEQUENCE 112 AA; 12339 MM; 7CFD38DBE8ED71 CRC64;

Query Match
Best Local Similarity 90.2%; Score 55; DB 1; Length 112;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PWTFGSGTKL 10
DB 99 PWTFGSGTKL 108

RESULT 10
KV3F_MOUSE STANDARD; PRT; 132 AA.
AC P01658;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region MOPC 321 precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-37.
RX MEDLINE=78235687; PubMed=98179;
RA Burstein Y., Schechter I.;
RT "Primary structures of N-terminal extra peptide segments linked to
RT the variable and constant regions of immunoglobulin light chain
RT precursors: implications on the organization and controlled
RT expression of immunoglobulin genes.";
RT Biochemistry 17:2392-2400(1978).
RN [2]
RP SEQUENCE OF 21-132.
RX MEDLINE=73140224; PubMed=4120629;
RA McKean D.J., Potter M., Hood L.E.;
RT "Mouse immunoglobulin chains. Partial amino acid sequence of a kappa
RT chain.";
RL Biochemistry 12:749-759(1973).
CC -I- MISCELLANEOUS: THE PARTIAL SEQUENCE OF THE C REGION OF THIS
CC BENCE-JONES PROTEIN WAS ALSO DETERMINED. IT DIFFERS FROM THAT
CC REPORTED FOR MOUSE MOPC 21 ONLY IN THE TRANSPOSITION OF TWO NEARBY
CC RESIDUES.

```


DR PIR: A01933; KWS32.
DR HSSP: P01607; IREI.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; Ig; 1.
KW Immunoglobulin V region; Bence-Jones protein; Signal.
FT SIGNAL 1 20
FT CHAIN 1 132 IG KAPPA CHAIN V-III REGION MOPC 321.
FT DOMAIN 21 43
FT DOMAIN 44 58
FT DOMAIN 59 73
FT DOMAIN 74 80
FT DOMAIN 81 112
FT DOMAIN 113 121
FT DOMAIN 122 131
FT DISULFID 43 112
FT NON_TER 132 132
SQ SEQUENCE 132 AA; 14523 MW; 9F3B809B773FBE9 CRC64;

Query Match 90.2%; Score 55; DB 1; Length 132;
Best Local Similarity 90.0%; Pred. No. 0.013;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 PWTGGGKTL 10
Db 119 PWTGGGKTL 128

RESULT 11
KV6A_MOUSE
ID KV6A_MOUSE STANDARD; PRT; 107 AA.
AC P01675;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-VI region XRPC 44.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79082830; Pubmed=103573;
RA Rao D.N., Rudikoff S., Potter M.;
RT "X Chain variable regions from three galactan binding myeloma
RT proteins.";
RL Biochemistry 17:5555-5559(1978).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT
CC BIND GALACTAN.
CC PIR: A01941; KWSX4.
DR HSSP: P01679; 2PBU.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; Ig; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 23
FT DOMAIN 24 33
FT DOMAIN 34 48
FT DOMAIN 49 55
FT DOMAIN 56 87
FT DOMAIN 88 96
FT DOMAIN 97 106
FT DISULFID 23 87
FT NON_TER 107 107
SQ SEQUENCE 107 AA; 11627 MW; 27A2D022BC0A34D7 CRC64;

Query Match 88.5%; Score 54; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 WTFGGGKTL 10
Db 95 WTFGGGKTL 103

RESULT 12
KV5Q_MOUSE
ID KV5Q_MOUSE STANDARD; PRT; 108 AA.
AC P01650;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-V region UPC 61.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79195288; Pubmed=109517;
RA Viana M., Rudikoff S., Potter M.;
RT "The structural basis of a hapten-inhibitable kappa-chain idiotype.";
RT J. Immunol. 122:1905-1910(1979).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT
CC BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).
CC PIR: A01939; KWS61.
DR HSSP: P80362; 1WTL.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; Ig; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 23
FT DOMAIN 24 34
FT DOMAIN 35 49
FT DOMAIN 50 56
FT DOMAIN 57 88
FT DOMAIN 89 97
FT DOMAIN 98 108
FT DISULFID 23 88
FT NON_TER 108 108
SQ SEQUENCE 108 AA; 11809 MW; FA64DA36076F2AFE CRC64;

Query Match 85.2%; Score 52; DB 1; Length 108;
Best Local Similarity 90.0%; Pred. No. 0.035;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGKTL 10
Db 95 PWTGGGKTL 104

RESULT 13
KV5R_MOUSE
ID KV5R_MOUSE STANDARD; PRT; 108 AA.
AC P01651;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-V region EPC 109.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79195288; Pubmed=109517;
RA Viana M., Rudikoff S., Potter M.;
RT "The structural basis of a hapten-inhibitable kappa-chain idiotype.";
RT J. Immunol. 122:1905-1910(1979).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT

—

```
Search completed: August 20, 2002, 13:17:53
Job time: 262 sec
```

Tue Aug 20 13:06:01 2002

us-09-824-286-4_copy_95_104.rsp

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:26 ; Search time 73.7 Seconds
(without alignments)
23.473 Million cell updates/sec

Title:	US-09-824-286-4_COPY_95_104
Perfect score:	61
Sequence:	1 PWTFGGCTKL 10

Scoring table: BLOSUM62

```
Searched:      562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222
```

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:
  Minimum Match 0%
  Maximum Match 100%
  Listing first 45 summaries

```

```
Database :

1:  sp_archaea:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mhc:*
8:  sp_oranelle:*
9:  sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	61	100.0	103	11	Q9JL80	Q9J180 mus musculus
2	61	100.0	107	11	Q9JL84	Q9J184 mus musculus
3	61	100.0	109	11	Q920E6	Q920E6 mus musculus
4	61	100.0	233	11	Q91WS9	Q91WS9 mus musculus
5	52	85.2	104	11	Q9JL82	Q9J182 mus musculus
6	52	85.2	111	11	Q920E9	Q920E9 mus musculus
7	52	85.2	127	11	Q925S9	Q925S9 mus musculus
8	52	85.2	298	11	Q9QYF0	Q9QYF0 mus musculus
9	50	82.0	114	4	Q9JL80	Q9J180 homo sapien
10	49	80.3	109	11	Q9ETJ3	Q9ETJ3 mus musculus
11	49	80.3	130	11	Q9DBM4	Q9DBM4 mus musculus
12	49	80.3	233	11	Q81V32	Q81V32 m adult mal
13	47	77.0	99	11	Q9JL74	Q9J174 mus musculus
14	47	77.0	211	11	Q9JLX0	Q9J1X0 mus musculus
15	46	75.4	238	11	Q99M37	Q99M37 mus musculus
16	45	73.8	97	11	Q9JL76	Q9J176 mus musculus

17	45	73.8	101	11	Q9JL78	Q9J178 mus musculus
18	45	73.8	106	5	Q9U410	Q9U410 schistosoma
19	45	73.8	107	11	Q9ER29	Q9ER29 mus musculus
20	45	73.8	109	4	Q9UL85	Q9UL85 homo sapien
21	45	73.8	109	4	Q9UL78	Q9UL78 homo sapien
22	45	73.8	214	11	Q9RLA5	Q9RLA5 mus musculus
23	45	73.8	234	11	Q9LWF8	Q9LWF8 mus musculus
24	43	70.5	108	4	Q9UL77	Q9UL77 homo sapien
25	43	70.5	241	11	Q921A6	Q921A6 mus musculus
26	43	70.5	618	5	Q62370	Q62370 caenorhabdi
27	41	67.2	638	16	Q9ZLC1	Q9ZLC1 helicobacte
28	41	67.2	651	16	Q9ZLB8	Q9ZLB8 helicobacte
29	41	67.2	847	4	Q96N13	Q96N13 homo sapien
30	41	67.2	936	16	Q9R2S3	Q9R2S3 delinococcus
31	40	65.6	107	4	Q96SA9	Q96SA9 homo sapien
32	40	65.6	108	4	Q9UL83	Q9UL83 homo sapien
33	40	65.6	116	4	Q96PF6	Q96PF6 homo sapien
34	40	65.6	185	10	Q946V1	Q946V1 zea mays (m
35	40	65.6	210	15	Q89986	Q89986 human immun
36	40	65.6	210	15	Q9YV97	Q9YV97 human immun
37	40	65.6	210	15	Q89997	Q89997 human immun
38	40	65.6	452	5	Q19175	Q19175 caenorhabdi
39	39	63.9	108	4	Q9UL70	Q9UL70 homo sapien
40	39	63.9	118	6	Q95J74	Q95J74 macaca fasc
41	39	63.9	367	12	Q9S7V1	Q9S7V1 chlamydomon
42	39	63.9	395	10	Q9NNA6	Q9NNA6 hyposoter d
43	39	63.9	414	17	Q9HOL6	Q9HOL6 halobacteri
44	39	63.9	460	8	Q79568	Q79568 musculus ma
45	39	63.9	460	8	Q9ZY36	Q9ZY36 raja radiat

ALIGNMENTS

RESULT	1			
Q9JL80	ID	Q9JL80	PRELIMINARY;	PRT; 103 AA.
AC	Q9JL80;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	ANTI-MYOSIN IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION (FRAGMENT).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCHI_TaxID-10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BALB/C;			
RX	MEDLINE-20448942; PubMed-10992488;			
RA	Malkiel S., Liao L., Cunningham M.W., Diamond B.;			
RT	"T-cell-dependent antibody response to the dominant epitope of streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive with cardiac myosin";			
RL	Infect. Immun. 68:5803-5808(2000).			
DR	EMBL; AF206026; AAF69324.1; -			
DR	HSSP; P80362; IWTL.			
DR	InterPro; IPR003006; IG_MHC.			
DR	InterPro; IPR003596; IG_v.			
DR	Pfam; PF00047; Ig_1.			
DR	SMART; SM00406; IgV_1.			
FT	NON_TER			
FT	1			
FT	103			
FT	103			
SEQUENCE	103 AA; 11224 MW; EC87D653DB3AAB21 CRC64;			

Query Match 100.0%; Score 61; DB 11; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.009;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

Db 91 PWTGCGTKL 100

RESULT 2

09JL84 PRELIMINARY; PRT: 107 AA.

AC 09JL84; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ANTI-MYOSIN IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION
DE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C;
RX MEDLINE=20448942; PubMed=10992488;
RA Malkiel S., Liao L., Cunningham M.W., Diamond B.;
RT "T-Cell-dependent antibody response to the dominant epitope of
streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive
with cardiac myosin."
FL Infect. Immun. 68:5803-5808(2000).
DR EMBL: AF206022; AAF69320.1; -.
DR HSSP: P80362; 1MTL.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig: 1.
DR SMART: SM00406; IGV: 1.
FT NON_TER 1
FT NON_TER 107
SQ SEQUENCE 107 AA; 11648 MW; ACF9B1253ACA1ESD CRC64;

Query Match

Best Local Similarity 100.0%; Score 61; DB 11; Length 107;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGCGTKL 10
Db 95 PWTGCGTKL 104

RESULT 3

0920E6 PRELIMINARY; PRT: 109 AA.

AC 0920E6; 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PTERIN-MIMICKING ANTI-IDIOPE KAPPA CHAIN VARIABLE REGION
DE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Atkin J.D., Iape A., Jennings I.G., Horaitis O., Cotton R.G.H.;
RT "Definition of the idiotope of Pterin-Mimicking Antibodies Expressed
in Mammalian Cells."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF307938; AAL09422.1; -.
FT NON_TER 1
FT NON_TER 109
SQ SEQUENCE 109 AA; 11943 MW; DAD3F98E05DD1501 CRC64;

Query Match 100.0%; Score 61; DB 11; Length 109;
Best Local Similarity 100.0%; Pred. No. 0.0096;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGCGTKL 10
Db 95 PWTGCGTKL 104

RESULT 4

091WS9 PRELIMINARY; PRT: 233 AA.

AC 091WS9; 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 25.8 KDA PROTEIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC013496; AAH13496.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 233 AA; 25781 MW; BIC184DA149A16EB CRC64;

Query Match

Best Local Similarity 100.0%; Score 61; DB 11; Length 233;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGCGTKL 10
Db 114 PWTGCGTKL 123

RESULT 5

09JL82 PRELIMINARY; PRT: 104 AA.

AC 09JL82; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ANTI-MYOSIN IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION
DE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C;
RX MEDLINE=20448942; PubMed=10992488;
RA Malkiel S., Liao L., Cunningham M.W., Diamond B.;
RT "T-Cell-dependent antibody response to the dominant epitope of
streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive
with cardiac myosin."
FL Infect. Immun. 68:5803-5808(2000).
DR EMBL: AF206024; AAF69322.1; -.
DR HSSP: P01607; 1REI.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig: 1.
DR SMART: SM00406; IGV: 1.
FT NON_TER 1
FT NON_TER 104
SQ SEQUENCE 104 AA; 11360 MW; 5DA8BBF05FOA1AE CRC64;

Query Match 85.2%; Score 52; DB 11; Length 104;
Best Local Similarity 90.0%; Pred. No. 0.27;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRKL 10
1:|||||
DB 92 PYTGGGTRKL 101

RESULT 6
ID Q920E9 PRELIMINARY; PRT: 111 AA.
AC Q920E9:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PTERIN-MIMICKING ANTI-IDIOPOPE KAPPA CHAIN VARIABLE REGION (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Atkin J.D., Tape A., Jennings I.G., Horaitis O., Cotton R.G.H.;
RT "Definition of the Idiotope of Pterin-Mimicking Antibodies Expressed in Mammalian Cells."
RL SUBMITTED (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF307935; AAL09419.1; -
FT NON_TER 1 111
FT SEQUENCE 111 AA; 12046 MW; 1E46988AA6858526 CRC64;

Query Match 85.2%; Score 52; DB 11; Length 111;
Best Local Similarity 90.0%; Pred. No. 0.29;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRKL 10
1:|||||
DB 99 PYTGGGTRKL 108

RESULT 7
ID Q925S9 PRELIMINARY; PRT: 127 AA.
AC Q925S9:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE IMMUNOGLOBULIN LIGHT CHAIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE=99306687; PubMed=10380019;
RA Tripathi P.K., Qin H., Bhattacharya-Chatterjee M., Ceriani R.L., Foon K.A., Chatterjee S.K.;
RT "Construction and characterization of a chimeric fusion protein consisting of an anti-idiotypic antibody mimicking a breast cancer-associated antigen and the cytokine GM-CSF."
RL Hydridoma 18:193-202(1999).
DR EMBL: AF144721; AAK5120.1; -
FT NON_TER 127 127
FT SEQUENCE 127 AA; 13794 MW; 13F61BBB981EAF5 CRC64;

Query Match 85.2%; Score 52; DB 11; Length 127;
Best Local Similarity 90.0%; Pred. No. 0.33;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRKL 10

DB 115 PYTGGGTRKL 124
1:|||||

RESULT 8
ID Q90YF0 PRELIMINARY; PRT: 298 AA.
AC Q90YF0:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CN 8 SCFV.
GN CN 8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=SPLEEN;
RX MEDLINE=20183931; PubMed=10706631;
RA Shiohara N., Demura T., Fukuda H.;
RT "Isolation of a vascular cell wall-specific monoclonal antibody recognizing a cell polarity by using a phage display subtraction method."
RL Proc. Natl. Acad. Sci. U.S.A. 97:2585-2590(2000).
DR EMBL: AB036341; BAA88633.1; -
DR HSP; P01607; IREL.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_2.
DR SMART: SM00406; IgV_2.
SQ SEQUENCE 298 AA; 31867 MW; E0F96B8A17004317 CRC64;

Query Match 85.2%; Score 52; DB 11; Length 298;
Best Local Similarity 90.0%; Pred. No. 0.79;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRKL 10
1:|||||
DB 267 PYTGGGTRKL 276

RESULT 9
ID Q90L80 PRELIMINARY; PRT: 114 AA.
AC Q90L80:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalls N.N., Berney S.M., Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal fetus."
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL: AF035034; AAD56270.1; -
DR HSP; P80362; IWTL.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; IgV_1.
FT NON_TER 114 114

```

SQ SEQUENCE 114 AA: 12775 MW: 070E31E210DICB01 CRC64;
Query Match
Best Local Similarity 80.0%; Score 50; DB 4; Length 114;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PWTGGGCTKL 10
    11111111
DB 101 PWTGGGCTKV 110

RESULT 10
C98T13 PRELIMINARY: PRT; 109 AA.
AC Q98T13;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE IMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RT "Nucleotide sequences encoding the variable regions of monoclonal AB,
RT A9-11-5, directed against antigen of Hepatitis B virus.";
RL Submitted (JUL-2000) to the EMBL/Genbank/DDb databases.
DR EMBL: AF287275; MAG3053.1; -.
DR HSPF; P01703; 7FAB.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_1.
DR SMART; SM00409; Ig_1.
DR SMART; SM00406; IgV_1.
DR SMART; SM00410; Ig_Like; 1.
FT NON_TER 1
FT TER 109
SQ SEQUENCE 109 AA: 11554 MW: 4F91E9D351B1E158 CRC64;

Query Match
Best Local Similarity 80.3%; Score 49; DB 11; Length 109;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTFGGCTKL 10
    11111111
DB 97 WTFGGCTKL 105

RESULT 11
C98D84 PRELIMINARY: PRT; 130 AA.
AC Q98D84;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE 1810027001R1K PROTEIN.
GN 1810027001R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shitagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
```

```

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kociba H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007622; BAB25142.1; -.
DR HSPF; P01842; 7FAB.
DR MGD; MGI:1916426; 1810027001R1K.
DR InterPro: IPR003597; Ig_C1.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IgC1_1.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
SQ SEQUENCE 130 AA: 14253 MW: 438197975E76E54 CRC64;

Query Match
Best Local Similarity 88.9%; Score 49; DB 11; Length 130;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTFGGCTKL 10
    11111111
DB 13 WTFGGCTKL 21

RESULT 12
C91V32 PRELIMINARY: PRT; 233 AA.
AC C91V32;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DE ADULT MALE SMALL INTESTINE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE CLONE:2010203H01, FULL INSERT SEQUENCE (ADULT MALE SMALL INTESTINE
DE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:2010005106, FULL
DE INSERT SEQUENCE) (ADULT MALE SMALL INTESTINE CDNA, RIKEN FULL-LENGTH
DE ENRICHED LIBRARY, CLONE:2010013G21, FULL INSERT SEQUENCE).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE=SMALL INTESTINE;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Aono H., Arai A.,
RA Arakawa T., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hizumoto K., Hirooka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kato H., Kawai J., Kojima Y.,
RA Kono H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Okazaki Y., Okido T.,
RA Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D.,
RA Shibata K., Shibata Y., Shitagawa A., Shiraki T., Sogabe Y.,
RA Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y.,
RA Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/Genbank/DDb databases.
(2)
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE=SMALL INTESTINE;
```


RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kitanai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishize T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Matsumi M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuda S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 DR EMBL: AK008421; BAB25659.1; -;
 DR EMBL: AK008129; BAB25479.1; -;
 DR EMBL: AK008232; BAB25546.1; -;
 SQ SEQUENCE 233 AA; 24874 MW; C4E896D46AABD2C1 CRC64;
 QY 2 WTEGCGTKL 10
 Db 116 WTEGCGTKL 124
 Query Match 80.3%; Score 49; DB 11; Length 233;
 Best Local Similarity 88.9%; Pred. No. 1.9;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 RESULT 13
 ID 09JL74 PRELIMINARY; PRT; 99 AA.
 AC 09JL74;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE ANTI-MYOSIN IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION
 (FRAGMENT).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RX MEDLINE=20448942; PubMed=10992488;
 RA Malkiel S., Liao L., Cunningham M.W., Diamond B.;
 RT "T-Cell dependent antibody response to the dominant epitope of
 streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive
 with cardiac myosin.";
 RL Infect. Immun. 68:5803-5808(2000).
 DR EMBL: AF206032; AAF69330.1; -;
 DR HSP; P80362; IWTL.

DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig; 1.
 DR SMART: SM00406; IGV; 1.
 FT NON_TER 1
 FT NON_TER 99
 SQ SEQUENCE 99 AA; 10939 MW; 3B25D0E78453324 CRC64;
 Query Match 77.0%; Score 47; DB 11; Length 99;
 Best Local Similarity 90.0%; Pred. No. 1.7;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PWTGCGTKL 10
 Db 87 PWTGCGTKL 96
 RESULT 14
 ID 09JXL0 PRELIMINARY; PRT; 211 AA.
 AC 09JXL0;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE ADULT MALE KIDNEY CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
 CD CLONE:0610010P20, FULL INSERT SEQUENCE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Aono H., Arai A.,
 RA Arikawa T., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kato H., Kawai J., Kojima Y.,
 RA Kono H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A.,
 RA Nishi K., Nomura K., Numazaki R., Ono M., Okazaki Y., Okido T.,
 RA Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D.,
 RA Shibata K., Shibata Y., Shinagawa A., Shiraki T., Sogabe Y.,
 RA Suzuki H., Tagami T., Tagawa A., Takahashi F., Tanaka T., Tejima Y.,
 RA Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
 RX MEDLINE=20530913; PubMed=11076861;

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
 RT "RIKEN Integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer."
 RL Genome Res. 10:1757-1771(2000).
 DR EMBL: AK002514; BAB22154.1; -
 SO SEQUENCE 211 AA; 23182 MW; 1A5FFA0F8BA50163 CRC64;

Query Match 77.0%; Score 47; DB 11; Length 211;
 Best Local Similarity 90.0%; Pred. No. 3.6;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PWTFGGGTKL 10
 1 | | | | | | | | | |
 Db 92 PRTFGGGTKL 101

RESULT 15
 Q99M37 PRELIMINARY; PRT; 238 AA.
 AC Q99M37;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHEtical 26.3 KDA PROTEIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD, GROSS
 RC TISSUE. ;
 RA Strausberg R.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC002035; AAH02035.1; -
 DR HSSP: P01679; 2PBJ.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003597; Ig_c1.
 DR InterPro: IPR003600; Ig_like.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003596; Ig_V.
 DR Pfam: PF00047; Ig_2.
 DR SMART: SM00409; Ig; 2.
 DR SMART: SM00407; IGc1; 1.
 DR SMART: SM00406; IGV; 1.
 DR SMART: SM00410; IG_like; 1.
 DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
 KW Hypothetical protein.
 SO SEQUENCE 238 AA; 26344 MW; FB2B06A0B801330A CRC64;

Query Match 75.4%; Score 46; DB 11; Length 238;
 Best Local Similarity 80.0%; Pred. No. 5.9;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PWTFGGGTKL 10
 1 | | | | | | | | | |
 Db 119 PYTFGGGTKL 128

Search completed: August 20, 2002, 13:17:27
 Job time: 241 sec

PT Blocking agents of the gamma common chain of cytokine receptors -
 PT particularly monoclonal antibodies, used to induce T cell anergy for
 PT treatment of immunological diseases
 XX
 PS Claim 24; Page 84; 111pp; English.
 CC
 CC This peptide comprises an epitope of the human cytokine receptor
 CC common gamma (gc) chain (see AAW31646) that is recognised by
 CC gc blocking agents of the invention. 5 Such epitopes (see
 CC AAW31650-54) have been identified. The invention provides
 CC compositions and methods for inhibiting cytokine signalling using
 CC gc chain blocking agents for the treatment of immunological
 CC diseases such as myasthenia gravis, rheumatoid arthritis, lupus,
 CC multiple sclerosis, insulin-dependent diabetes, inflammatory bowel
 CC disease, sympathetic ophthalmia, uveitis, allergy, asthma,
 CC parasitic infection, graft vs. host disease or psoriasis. A
 CC preferred gc blocking agent is Mab CP-B8 or its Fab fragment (see
 CC also AAW31647-48).
 CC
 SQ Sequence 5 AA:

Query Match 100.0%; Score 28; DB 19; Length 5;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
 Db 1 fnvey 5

RESULT 2
 AAB48710
 ID AAB48710 standard; peptide; 13 AA.

XX AAB48710;

XX 09-MAR-2001 (first entry)

DE Human cytokine receptor subunit h-gamma A-B loop.

XX
 XX Haematopoietic growth factor receptor; cytokine receptor; subunit;
 KM CR domain; CRD; h-beta-c subunit; GM-CSF receptor; IL-3 receptor;
 KM IL-5 receptor; agonist; antagonist; myeloid leukemia; cancer;
 KM lymphocyte leukaemia; non-haematopoietic tumour; inflammatory disease;
 KM asthma; rheumatoid arthritis; atherosclerosis; CR module; CRM; A-B loop.
 XX
 OS Homo sapiens.

XX WO200066632-A1.

PD 09-NOV-2000.

PE 01-MAY-2000; 2000WO-AU00394.

PR 29-APR-1999; 99AU-0000053.

PA (MEDV-) MEDVET SCI PTY LTD.

PI D'Andrea R, Bagley C, Vadas MA;

XX WPI; 2001-049688/06.

DR
 XX
 XX New agonists or antagonists of haematopoietic growth factors for treating
 PT myeloid and lymphocyte leukemias, tumors and acute and chronic
 PT inflammation such as asthma, rheumatoid arthritis and atherosclerosis -
 XX
 PS Disclosure; Fig 2; 35pp; English.

XX The invention relates to a novel agonist or antagonist of a haematopoietic
 CC growth factor which is capable of binding a region of the CRD3 (cytokine
 CC receptor domain 3) of the human GM-CSF receptor/IL-3R/IL-5R shared
 CC subunit h-beta-c, or an equivalent CRD in a different haematopoietic

CC growth factor receptor or other cytokine receptor. Binding of the
 CC agonist or antagonist has an effect on the interaction between CRD3 and
 CC CRD4 (or analogous domain) which mediates the agonist or antagonist
 CC property. In particular, the agonist or antagonist binds to the A-B loop
 CC or the E-F loop of CRD3 or its equivalent. The invention also relates to
 CC a method for screening and isolating a haematopoietic growth factor agonist
 CC or antagonist, and pharmaceutical compositions containing an agonist or
 CC antagonist. The agonists are used for treating conditions currently
 CC treated by GM-CSF (granulocyte macrophage-colony stimulating factor),
 CC IL-3 (interleukin-3), IL-5, and other members of the family of
 CC haematopoietic growth factors. Haematopoietic growth factor antagonists are
 CC useful for treating diseases such as myeloid and lymphocyte leukemias,
 CC tumours of non-haematopoietic origin and acute and chronic inflammatory
 CC conditions (e.g., asthma, rheumatoid arthritis and atherosclerosis).
 CC Sequences AAB48693-B48716 represent the A-B loops of a variety of
 CC cytokine receptor subunits from human and mouse.
 CC
 SQ Sequence 13 AA:

Query Match 100.0%; Score 28; DB 22; Length 13;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
 Db 4 fnvey 8

RESULT 3
 AAB23875
 ID AAB23875 standard; peptide; 21 AA.

XX AAB23875;

XX 17-JAN-2001 (first entry)

DE Artificial sequence designed peptide #1.

XX Peptide synthesis; chemical synthesis; solid phase synthesis.

XX Synthetic.

PN WO200055182-A1.

PD 21-SEP-2000.

PE 15-MAR-2000; 2000WO-JP01584.

PR 15-MAR-1999; 99JP-0067917.

PA (SANY) SANKYO CO LTD.

PI Kawaguchi J, Serizawa N;

DR WPI; 2000-602106/57.

XX Chemically synthesizing a peptide by solid phase synthesis by adding
 PT protective group to inactivate part of alpha amino group of an amino
 PT acid group -
 XX

PS Example 1; Page 8; 38pp; Japanese.

XX The present invention describes a method for chemically synthesizing a
 CC peptide by solid phase synthesis. The method comprises processing a
 CC carrier resin to which the C-terminal amino acid derivative of the
 CC target peptide has been preliminarily bonded is characterised by adding
 CC a protective group to inactivate a part of alpha-amino groups of an
 CC amino acid capable of undergoing a peptide extension reaction on the
 CC carrier resin. The method is useful for synthesizing a peptide by solid
 CC phase synthesis. Long chain peptides can be readily synthesised in with
 CC a high success ratio without using the segment condensation method or
 CC gene manipulations. The present sequence represents an amino acid

CC peptide sequence given in an example from the present invention.
XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 28; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FNVEY 5
|||||
DB 5 fnvey 9

RESULT 4
AAB23886
ID AAB23886 standard; peptide; 21 AA.
XX
AC AAB23886;

DT 17-JAN-2001 (first entry)
XX
DE Artificial sequence designed peptide #10.
XX
KW Peptide synthesis; chemical synthesis; solid phase synthesis.
XX
OS Synthetic.

XX
PN WO200055182-A1.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-JP01584.

XX
PR 15-MAR-1999; 99JP-0067917.
XX
PA (SANY) SANKYO CO LTD.
XX
PI Kawaguchi J, Serizawa N;
XX
DR WPI; 2000-602106/57.

PT Chemically synthesizing a peptide by solid phase synthesis by adding
PT protective group to inactivate part of alpha amino group of an amino
PT acid group
XX
PS Example 3; Page 15; 38pp; Japanese.

CC The present invention describes a method for chemically synthesizing a
CC peptide by solid phase synthesis. The method comprises processing a
CC carrier resin to which the C-terminal amino acid derivative of the
CC target peptide has been preliminarily bonded is characterised by adding
CC a protective group to inactivate a part of alpha-amino groups of an
CC amino acid capable of undergoing a peptide extension reaction on the
CC carrier resin. The method is useful for synthesising a peptide by solid
CC phase synthesis. Long chain peptides can be readily synthesised in with
CC a high success ratio without using the segment condensation method or
CC gene manipulations. The present sequence represents an amino acid
CC peptide sequence given in an example from the present invention.

XX
SQ Sequence 21 AA;

Query Match 100.0%; Score 28; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
|||||
DB 5 fnvey 9

RESULT 5

AAB23876
ID AAB23876 standard; peptide; 31 AA.
XX
AC AAB23876;

DT 17-JAN-2001 (first entry)
XX
DE Artificial sequence designed peptide #2.
XX
KW Peptide synthesis; chemical synthesis; solid phase synthesis.

XX
OS Synthetic.
XX
PN WO200055182-A1.
XX
PD 21-SEP-2000.

XX
PF 15-MAR-2000; 2000WO-JP01584.
XX
PR 15-MAR-1999; 99JP-0067917.

XX
PA (SANY) SANKYO CO LTD.
XX
PI Kawaguchi J, Serizawa N;
XX
DR WPI; 2000-602106/57.

PT Chemically synthesizing a peptide by solid phase synthesis by adding
PT protective group to inactivate part of alpha amino group of an amino
PT acid group
XX
PS Example 1; Page 9; 38pp; Japanese.

CC The present invention describes a method for chemically synthesizing a
CC peptide by solid phase synthesis. The method comprises processing a
CC carrier resin to which the C-terminal amino acid derivative of the
CC target peptide has been preliminarily bonded is characterised by adding
CC a protective group to inactivate a part of alpha-amino groups of an
CC amino acid capable of undergoing a peptide extension reaction on the
CC carrier resin. The method is useful for synthesising a peptide by solid
CC phase synthesis. Long chain peptides can be readily synthesised in with
CC a high success ratio without using the segment condensation method or
CC gene manipulations. The present sequence represents an amino acid
CC peptide sequence given in an example from the present invention.

XX
SQ Sequence 31 AA;

Query Match 100.0%; Score 28; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
|||||
DB 15 fnvey 19

RESULT 6
AAB23887
ID AAB23887 standard; peptide; 41 AA.
XX
AC AAB23887;

DT 17-JAN-2001 (first entry)
XX
DE Artificial sequence designed peptide #11.
XX
KW Peptide synthesis; chemical synthesis; solid phase synthesis.

XX
OS Synthetic.
XX
PN WO200055182-A1.

XX

PD 21-SEP-2000.
XX
PS 15-MAR-2000; 2000WO-JP01584.
XX
PR 15-MAR-1999; 99JP-0067917.
XX
PA (SANY) SANKYO CO LTD.
XX
PI Kawaguchi J, Serizawa N;
XX
DR WPI; 2000-602106/57.
XX
PT Chemically synthesizing a peptide by solid phase synthesis by adding
PT protective group to inactivate part of alpha amino group of an amino
PT acid group
XX
PS Example 3; Page 15-16; 38pp; Japanese.
XX
CC The present invention describes a method for chemically synthesizing a
CC peptide by solid phase synthesis. The method comprises processing a
CC carrier resin to which the C-terminal amino acid derivative of the
CC target peptide has been preliminarily bonded is characterised by adding
CC a protective group to inactivate a part of alpha-amino groups of an
CC amino acid capable of undergoing a peptide extension reaction on the
CC carrier resin. The method is useful for synthesising a peptide by solid
CC phase synthesis. Long chain peptides can be readily synthesised in with
CC a high success ratio without using the segment condensation method or
CC gene manipulations. The present sequence represents an amino acid
CC peptide sequence given in an example from the present invention.
XX
SQ Sequence 41 AA;

Query Match 100.0%; Score 28; DB 21; Length 41;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
| | | | |
Db 25 fnvey 29

RESULT 7
AAB23877
ID AAB23877 standard; peptide: 51 AA.
XX
AC AAB23877;
XX
DT 17-JAN-2001 (first entry)
XX
DE Artificial sequence designed peptide #3.
XX
KM Peptide synthesis; chemical synthesis; solid phase synthesis.
XX
OS Synthetic.
XX
PN WO200055182-A1.
XX
PD 21-SEP-2000.
XX
PS 15-MAR-2000; 2000WO-JP01584.
XX
PR 15-MAR-1999; 99JP-0067917.
XX
PA (SANY) SANKYO CO LTD.
XX
PI Kawaguchi J, Serizawa N;
XX
DR WPI; 2000-602106/57.
XX
PT Chemically synthesizing a peptide by solid phase synthesis by adding
PT protective group to inactivate part of alpha amino group of an amino
PT acid group

XX
PS Example 1; Page 9; 38pp; Japanese.
XX
CC The present invention describes a method for chemically synthesizing a
CC peptide by solid phase synthesis. The method comprises processing a
CC carrier resin to which the C-terminal amino acid derivative of the
CC target peptide has been preliminarily bonded is characterised by adding
CC a protective group to inactivate a part of alpha-amino groups of an
CC amino acid capable of undergoing a peptide extension reaction on the
CC carrier resin. The method is useful for synthesising a peptide by solid
CC phase synthesis. Long chain peptides can be readily synthesised in with
CC a high success ratio without using the segment condensation method or
CC gene manipulations. The present sequence represents an amino acid
CC peptide sequence given in an example from the present invention.
XX
SQ Sequence 51 AA;

Query Match 100.0%; Score 28; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
| | | | |
Db 35 fnvey 39

RESULT 8
AAB23888
ID AAB23888 standard; peptide: 51 AA.
XX
AC AAB23888;
XX
DT 17-JAN-2001 (first entry)
XX
DE Artificial sequence designed peptide #12.
XX
KM Peptide synthesis; chemical synthesis; solid phase synthesis.
XX
OS Synthetic.
XX
PN WO200055182-A1.
XX
PD 21-SEP-2000.
XX
PS 15-MAR-2000; 2000WO-JP01584.
XX
PR 15-MAR-1999; 99JP-0067917.
XX
PA (SANY) SANKYO CO LTD.
XX
PI Kawaguchi J, Serizawa N;
XX
DR WPI; 2000-602106/57.
XX
PT Chemically synthesizing a peptide by solid phase synthesis by adding
PT protective group to inactivate part of alpha amino group of an amino
PT acid group
XX
PS Example 3; Page 16; 38pp; Japanese.
XX
CC The present invention describes a method for chemically synthesizing a
CC peptide by solid phase synthesis. The method comprises processing a
CC carrier resin to which the C-terminal amino acid derivative of the
CC target peptide has been preliminarily bonded is characterised by adding
CC a protective group to inactivate a part of alpha-amino groups of an
CC amino acid capable of undergoing a peptide extension reaction on the
CC carrier resin. The method is useful for synthesising a peptide by solid
CC phase synthesis. Long chain peptides can be readily synthesised in with
CC a high success ratio without using the segment condensation method or
CC gene manipulations. The present sequence represents an amino acid
CC peptide sequence given in an example from the present invention.
XX

SQ Sequence 51 AA;

Query Match 100.0%; Score 28; DB 21; Length 51;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
 |||||
 Db 35 fnvey 39

RESULT 9

AAB23889
 ID AAB23889 standard; peptide; 61 AA.

AC AAB23889;

DT 17-JAN-2001 (first entry)

DE Artificial sequence designed peptide #13.

KW Peptide synthesis; chemical synthesis; solid phase synthesis.

XX Synthetic.

OS WO200055182-A1.

PN 21-SEP-2000.

PD 15-MAR-2000; 2000WO-JP01584.

PF 15-MAR-1999; 99JP-0067917.

PR (SANY) SANKYO CO LTD.

PA Kawaguchi J, Serizawa N;

PI WPI; 2000-602106/57.

DR Chemically synthesizing a peptide by solid phase synthesis by adding

PT protective group to inactivate part of alpha amino group of an amino

PT acid group

XX Example 3; Page 16; 38pp; Japanese.

CC The present invention describes a method for chemically synthesizing a
 CC peptide by solid phase synthesis. The method comprises processing a
 CC carrier resin to which the C-terminal amino acid derivative of the
 CC target peptide has been preliminarily bonded is characterised by adding
 CC a protective group to inactivate a part of alpha-amino groups of an
 CC amino acid capable of undergoing a peptide extension reaction on the
 CC carrier resin. The method is useful for synthesising a peptide by solid
 CC phase synthesis. Long chain peptides can be readily synthesised in with
 CC a high success ratio without using the segment condensation method or
 CC gene manipulations. The present sequence represents an amino acid
 CC peptide sequence given in an example from the present invention.

SQ Sequence 61 AA;

Query Match 100.0%; Score 28; DB 21; Length 61;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
 |||||
 Db 45 fnvey 49

RESULT 10

AAB23878
 ID AAB23878 standard; peptide; 71 AA.

XX AAB23878;
 AC

XX 17-JAN-2001 (first entry)
 DT

XX Artificial sequence designed peptide #4.
 DE

XX Peptide synthesis; chemical synthesis; solid phase synthesis.
 KW

XX Synthetic.
 OS

XX WO200055182-A1.
 PN

XX 21-SEP-2000.
 PD

XX 15-MAR-2000; 2000WO-JP01584.
 PF

XX 15-MAR-1999; 99JP-0067917.
 PR

XX (SANY) SANKYO CO LTD.
 PA

XX Kawaguchi J, Serizawa N;
 PI

XX WPI; 2000-602106/57.
 DR

XX Chemically synthesizing a peptide by solid phase synthesis by adding
 PT protective group to inactivate part of alpha amino group of an amino
 PT acid group

XX Example 1; Page 10; 38pp; Japanese.
 PS

CC The present invention describes a method for chemically synthesizing a
 CC peptide by solid phase synthesis. The method comprises processing a
 CC carrier resin to which the C-terminal amino acid derivative of the
 CC target peptide has been preliminarily bonded is characterised by adding
 CC a protective group to inactivate a part of alpha-amino groups of an
 CC amino acid capable of undergoing a peptide extension reaction on the
 CC carrier resin. The method is useful for synthesising a peptide by solid
 CC phase synthesis. Long chain peptides can be readily synthesised in with
 CC a high success ratio without using the segment condensation method or
 CC gene manipulations. The present sequence represents an amino acid
 CC peptide sequence given in an example from the present invention.

SQ Sequence 71 AA;

Query Match 100.0%; Score 28; DB 21; Length 71;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
 |||||
 Db 55 fnvey 59

RESULT 11

AAB23885
 ID AAB23885 standard; peptide; 79 AA.

AC AAB23885;

DT 17-JAN-2001 (first entry)

DE Artificial sequence designed peptide SEQ ID NO:3.

KW Peptide synthesis; chemical synthesis; solid phase synthesis.

XX Synthetic.

PN WO200055182-A1.

PD 21-SEP-2000.

PF 15-MAR-2000; 2000MO-JP01584.
 XX
 PR 15-MAR-1999; 99JP-0067917.
 XX
 PA (SANY) SANKYO CO LTD.
 XX
 PI Kawaguchi J, Serizawa N;
 XX
 DR WPI; 2000-602106/57.
 XX
 PT Chemically synthesizing a peptide by solid phase synthesis by adding
 PT protective group to inactivate part of alpha amino group of an amino
 PT acid group
 XX
 PS Example 3; Page 32-33; 38pp; Japanese.
 XX
 CC The present invention describes a method for chemically synthesizing a
 CC peptide by solid phase synthesis. The method comprises processing a
 CC carrier resin to which the C-terminal amino acid derivative of the
 CC target peptide has been preliminarily bonded is characterised by adding
 CC a protective group to inactivate a part of alpha-amino groups of an
 CC amino acid capable of undergoing a peptide extension reaction on the
 CC carrier resin. The method is useful for synthesizing a peptide by solid
 CC phase synthesis. Long chain peptides can be readily synthesised in with
 CC a high success ratio without using the segment condensation method or
 CC gene manipulations. The present sequence represents an amino acid
 CC peptide sequence given in an example from the present invention.
 XX
 SQ Sequence 79 AA;

Query Match 100.0%; Score 28; DB 21; Length 79;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 FNVEY 5
 |||||
 Db 63 fnvey 67

RESULT 12
 AAB23890
 ID AAB23890 standard; peptide; 79 AA.
 XX
 AC AAB23890;
 XX
 DT 17-JAN-2001 (first entry)
 XX
 DE Artificial sequence designed peptide #14.
 XX
 KW Peptide synthesis; chemical synthesis; solid phase synthesis.
 XX
 OS Synthetic.
 XX
 PN WO200055182-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 15-MAR-2000; 2000MO-JP01584.
 XX
 PR 15-MAR-1999; 99JP-0067917.
 XX
 PA (SANY) SANKYO CO LTD.
 XX
 PI Kawaguchi J, Serizawa N;
 XX
 DR WPI; 2000-602106/57.
 XX
 PT Chemically synthesizing a peptide by solid phase synthesis by adding
 PT protective group to inactivate part of alpha amino group of an amino
 PT acid group
 XX
 PS Example 3; Page 17; 38pp; Japanese.

XX
 CC The present invention describes a method for chemically synthesizing a
 CC peptide by solid phase synthesis. The method comprises processing a
 CC carrier resin to which the C-terminal amino acid derivative of the
 CC target peptide has been preliminarily bonded is characterised by adding
 CC a protective group to inactivate a part of alpha-amino groups of an
 CC amino acid capable of undergoing a peptide extension reaction on the
 CC carrier resin. The method is useful for synthesizing a peptide by solid
 CC phase synthesis. Long chain peptides can be readily synthesised in with
 CC a high success ratio without using the segment condensation method or
 CC gene manipulations. The present sequence represents an amino acid
 CC peptide sequence given in an example from the present invention.
 XX
 SQ Sequence 79 AA;

Query Match 100.0%; Score 28; DB 21; Length 79;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 FNVEY 5
 |||||
 Db 63 fnvey 67

RESULT 13
 AAB69096
 ID AAB69096 standard; Peptide; 79 AA.
 XX
 AC AAB69096;
 XX
 DT 23-APR-2001 (first entry)
 XX
 DE Cedar pollen allergen T cell epitope derived peptide SEQ ID NO:3.
 XX
 KW Japanese cedar; Cryptomeria japonica; cedar pollen allergen;
 KW T cell epitope; antisuigipollinosis.
 XX
 OS Cryptomeria japonica.
 XX
 PN JP2000327699-A.
 XX
 PD 28-NOV-2000.
 XX
 PF 15-MAR-2000; 2000JP-0071710.
 XX
 PR 15-MAR-1999; 99JP-0068316.
 XX
 PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 XX
 PI (SANY) SANKYO CO LTD.
 XX
 DR WPI; 2001-185061/19.
 XX
 PT Novel peptide and its use -
 XX
 PS Claim 13; Page 32-33; 75pp; Japanese.
 XX
 CC The present invention describes a peptide, its complex, derivative or
 CC its polymerizate, where the peptide (I) has a formula of:
 CC alpha1-alpha2-alpha3-alpha4-alpha5-alpha6-alpha7
 CC where alpha1 to alpha7 = amino acid sequence selected from amino acid
 CC sequences ranging from 11-19 amino acids derived from T cell epitopes
 CC derived from cedar (Japanese cedar - Cryptomeria japonica) pollen
 CC allergens. The peptide can be used in an antisuigipollinosis agent.
 CC AAF59011 to AAF59062 and AAB69094 to AAB69121 represent sequences used
 CC in the exemplification of the present invention.
 XX
 SQ Sequence 79 AA;

Query Match 100.0%; Score 28; DB 22; Length 79;
 Best Local Similarity 100.0%; Pred. No. 15;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5
|||||

Db 63 fnvey 67

RESULT 14

AAB23874
ID AAB23874 standard; peptide; 81 AA.

AC AAB23874;

DT 17-JAN-2001 (first entry)

DE Artificial sequence designed peptide SEQ ID NO:1.

KW Peptide synthesis; chemical synthesis; solid phase synthesis.

OS Synthetic.

PN WO200055182-A1.

PD 21-SEP-2000.

PF 15-MAR-2000; 2000WO-JP01584.

PR 15-MAR-1999; 99JP-0067917.

PA (SANY) SANKYO CO LTD.

PI Kawaguchi J, Serizawa N;

DR WPI; 2000-602106/57.

PT Chemically synthesizing a peptide by solid phase synthesis by adding
protective group to inactivate part of alpha amino group of an amino
acid group

PS Example 1; Page 31; 38pp; Japanese.

CC The present invention describes a method for chemically synthesizing a
CC peptide by solid phase synthesis. The method comprises processing a
CC carrier resin to which the C-terminal amino acid derivative of the
CC target peptide has been preliminarily bonded is characterised by adding
CC a protective group to inactivate a part of alpha-amino groups of an
CC amino acid capable of undergoing a peptide extension reaction on the
CC carrier resin. The method is useful for synthesising a peptide by solid
CC phase synthesis. Long chain peptides can be readily synthesised in with
CC a high success ratio without using the segment condensation method or
CC gene manipulations. The present sequence represents an amino acid
CC peptide sequence given in an example from the present invention.

XX Sequence 81 AA;

Query Match 100.0%; Score 28; DB 21; Length 81;

Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5
|||||

Db 65 fnvey 69

RESULT 15

AAB23879
ID AAB23879 standard; peptide; 81 AA.

AC AAB23879;

DT 17-JAN-2001 (first entry)

DE Artificial sequence designed peptide #5.

KW Peptide synthesis; chemical synthesis; solid phase synthesis.

OS Synthetic.

PN WO200055182-A1.

PD 21-SEP-2000.

PF 15-MAR-2000; 2000WO-JP01584.

PR 15-MAR-1999; 99JP-0067917.

PA (SANY) SANKYO CO LTD.

PI Kawaguchi J, Serizawa N;

DR WPI; 2000-602106/57.

PT Chemically synthesizing a peptide by solid phase synthesis by adding
protective group to inactivate part of alpha amino group of an amino
acid group

PS Example 1; Page 10; 38pp; Japanese.

CC The present invention describes a method for chemically synthesizing a
CC peptide by solid phase synthesis. The method comprises processing a
CC carrier resin to which the C-terminal amino acid derivative of the
CC target peptide has been preliminarily bonded is characterised by adding
CC a protective group to inactivate a part of alpha-amino groups of an
CC amino acid capable of undergoing a peptide extension reaction on the
CC carrier resin. The method is useful for synthesising a peptide by solid
CC phase synthesis. Long chain peptides can be readily synthesised in with
CC a high success ratio without using the segment condensation method or
CC gene manipulations. The present sequence represents an amino acid
CC peptide sequence given in an example from the present invention.

XX Sequence 81 AA;

Query Match 100.0%; Score 28; DB 21; Length 81;

Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5
|||||

Db 65 fnvey 69

Search completed: August 20, 2002, 11:07:15
Job time: 6100 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:10:33 ; Search time 100.84 Seconds
(without alignments)
4.764 Million cell updates/sec

Title: US-09-824-286-13

Perfect score: 28

Sequence: 1 FNVEY 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR71:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	149	2 D97218	probable membrane
2	28	100.0	176	2 S28944	ribosomal protein
3	28	100.0	176	2 S55970	ribosomal protein
4	28	100.0	186	2 A81375	GTP cyclonhydrolase
5	28	100.0	259	1 CRB02	carbonate dehydrat
6	28	100.0	276	1 B64372	ferredoxin - Metha
7	28	100.0	299	2 T11070	NADH dehydrogenase
8	28	100.0	301	2 G96944	probable permenase
9	28	100.0	308	2 S58995	NADH dehydrogenase
10	28	100.0	312	2 S01191	NADH dehydrogenase
11	28	100.0	313	2 B71390	NADH dehydrogenase
12	28	100.0	314	2 T12010	NADH dehydrogenase
13	28	100.0	314	2 T09813	NADH dehydrogenase
14	28	100.0	315	1 OXMSIM	NADH dehydrogenase
15	28	100.0	317	2 T11337	NADH dehydrogenase
16	28	100.0	318	1 OXB01M	NADH dehydrogenase
17	28	100.0	318	1 OQRTIM	NADH dehydrogenase
18	28	100.0	318	2 A58850	NADH dehydrogenase
19	28	100.0	318	2 A58850	NADH dehydrogenase
20	28	100.0	318	2 S41835	NADH dehydrogenase
21	28	100.0	318	2 S47870	NADH dehydrogenase
22	28	100.0	318	2 S26151	NADH dehydrogenase
23	28	100.0	318	2 S41820	NADH dehydrogenase
24	28	100.0	318	2 A58888	NADH dehydrogenase
25	28	100.0	318	2 T11441	NADH dehydrogenase
26	28	100.0	318	2 T11493	NADH dehydrogenase
27	28	100.0	318	2 T11389	NADH dehydrogenase
28	28	100.0	318	2 T11857	NADH dehydrogenase
29	28	100.0	318	2 T10972	NADH dehydrogenase
				2 T11050	NADH dehydrogenase

30	28	100.0	318	2 T11480	NADH dehydrogenase
31	28	100.0	318	2 T11140	NADH dehydrogenase
32	28	100.0	318	2 T11247	NADH dehydrogenase
33	28	100.0	318	2 T11363	NADH dehydrogenase
34	28	100.0	318	2 T11428	NADH dehydrogenase
35	28	100.0	318	2 T45550	NADH dehydrogenase
36	28	100.0	319	2 T11454	NADH dehydrogenase
37	28	100.0	321	2 S55004	NADH dehydrogenase
38	28	100.0	321	2 T11274	NADH dehydrogenase
39	28	100.0	321	2 T13811	NADH dehydrogenase
40	28	100.0	322	2 S68128	NADH dehydrogenase
41	28	100.0	322	2 A90620	NADH dehydrogenase
42	28	100.0	322	2 A99626	NADH dehydrogenase
43	28	100.0	322	2 T11101	NADH dehydrogenase
44	28	100.0	322	2 T11127	NADH dehydrogenase
45	28	100.0	323	1 QXXLIM	NADH dehydrogenase

ALIGNMENTS

```
RESULT 1
D97218
probable membrane protein CAC2586 [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: D97218
R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; I
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: D97218
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-149 <KUR>
A:Cross-References: GB:AE001437; PIDN:AKR80535.1; PID:915025610; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2586

Query Match      100.0%; Score 28; DB 2; Length 149;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
Db 67 FNVEY 71

RESULT 2
S28944
ribosomal protein L6.e.A, cytosolic - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YML073c; ribosomal protein YL16.A
C:Species: Saccharomyces cerevisiae
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
C:Accession: S28944; S78086
R:Hashimoto, T.; Suzuki, K.; Mizuta, K.; Otaka, E.
Biochim. Biophys. Acta 1132, 195-198, 1992
A:Title: Yeast ribosomal proteins: XIV. Complete nucleotide sequences of the two gene
A:Reference number: S28944; MUID:93003323
A:Accession: S28944
A:Molecule type: DNA
A:Residues: 1-176 <HAS>
A:Cross-References: EMBL:D10225; NID:g218508; PIDN:BAA01077.1; PID:g218509
R:Brown, D.; Bowman, S.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48816
A:Accession: S78086
A:Molecule type: DNA
A:Residues: 1-176 <BRO>
A:Cross-References: EMBL:D46373; NID:g587529; PID:g914877; MIPS:YML073c
```

C:Genetics:
 A:Gene: SGD:YIL16A
 A:Cross-references: MIPS:YML073c; SGD:S0004538
 A:Map position: 13L
 A:Introns: 5/3
 A:Note: YML073c
 C:Keywords: cytosol; protein biosynthesis; ribosome

Query Match 100.0%; Score 28; DB 2; Length 176;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
 Db 101 FNVEY 105

RESULT 3
 S53970
 ribosomal protein L6.e.B, cytosolic - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein L9324.4; protein YLR448w; ribosomal protein YLR16.B
 C:Species: Saccharomyces cerevisiae
 C:Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 21-Jul-2000
 C:Accession: S53970; S28945; S11257
 R:Du, Z.
 submitted to the EMBL Data Library, March 1995
 A:Description: The sequence of S. cerevisiae cosmid 9324.
 A:Reference number: S53966
 A:Accession: S53970
 A:Molecule type: DNA
 A:Residues: 1-176 <DOZ>
 A:Cross-references: GB:022382; NID:g717059; PID:g717063; MIPS:YLR448w
 A:Experimental sources: strain S288C (AB972)
 R:Hashimoto, T.; Suzuki, K.; Mizuta, K.; Otake, E.
 Biochim. Biophys. Acta 1132, 195-198, 1992
 A:Title: Yeast ribosomal proteins: XIV. Complete nucleotide sequences of the two genes
 A:Reference number: S28944; MUID:93003323
 A:Accession: S28945
 A:Molecule type: DNA
 A:Residues: 1-67, 'L', '69-176 <HNS>
 A:Cross-references: EMBL:D10226; NID:g218510; PIDN:BA01078.1; PID:g218511
 R:Otake, E.; Higo, K.I.; Itoh, T.
 Mol. Gen. Genet. 195, 544-546, 1984
 A:Title: Yeast ribosomal proteins. VIII. Isolation of two proteins and sequence character
 A:Reference number: S11249
 A:Accession: S11257
 A:Molecule type: protein
 A:Residues: 2-8, 'X', '10-12, '2B', '15-25 <OTR>
 C:Genetics:
 A:Gene: SGD:RPL16B; YLR16B
 A:Cross-references: SGD:S0004440; MIPS:YLR448w
 A:Map position: 12R
 A:Introns: 5/3
 C:Keywords: cytosol; protein biosynthesis; ribosome
 F:2-176/Product: ribosomal protein L6.e #status experimental <MAT>

Query Match 100.0%; Score 28; DB 2; Length 176;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
 Db 101 FNVEY 105

RESULT 4
 AB1375
 GTP cyclohydrolase II (EC 3.5.4.25) Cj0996 [imported] - Campylobacter jejuni (strain NCT
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
 C:Accession: AB1375

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chl
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Ba
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
 A:Reference number: AB1250; MUID:20150912
 A:Accession: AB1375
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-186 <PAR>
 A:Cross-references: GB:AL139076; GB:AL111168; NID:96968128; PIDN:CAB73252.1; PID:9696
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: rba; Cj0996
 C:Superfamily: Escherichia coli cyclohydrolase II; cyclohydrolase homology
 C:Keywords: hydrolase

Query Match 100.0%; Score 28; DB 2; Length 186;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
 Db 170 FNVEY 174

RESULT 5
 CR802
 carbonate dehydratase (EC 4.2.1.1) II - bovine (tentative sequence)
 N:Alternate names: carbonic anhydrase II
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 31-Mar-1981 #sequence_revision 31-Mar-1981 #text_change 31-Mar-2000
 C:Accession: A01144
 R:Sciaky, M.; Limozin, N.; Filippi-Foveau, D.; Guillan, J.M.; Laurent-Tabasse, G.
 Biochimie 58, 1071-1082, 1976
 A:Title: Structure primaire de l'anhydrase carbonique erythrocytaire bovine CI. II.
 A:Reference number: A90669; MUID:77065798
 A:Accession: A01144
 A:Molecule type: protein
 A:Residues: 1-259 <SCI>
 R:Guilan, J.M.; Limozin, N.; Mallet, B.; Di Costanzo, J.; Charrel, M.
 Biochimie 59, 293-302, 1977
 A:Title: Independance genetique de deux formes de l'anhydrase carbonique erythrocytal
 A:Reference number: A90672; MUID:77242599
 A:Contents: annotation
 A:Note: one minor and two major forms were isolated chromatographically. One of the m
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology
 C:Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc
 F:4-258/Domain: carbonic anhydrase homology <CAH>
 F:1/Modified site: acetylated amino end (Ser) #status experimental
 F:93,95,118/Binding site: zinc (His) #status predicted

Query Match 100.0%; Score 28; DB 1; Length 259;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
 Db 65 FNVEY 69

RESULT 6
 B64372
 ferredoxin - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: B64372
 R:Butt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
 R:Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodex,
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*
A:Reference number: A64300; MUID:96337999
A:Accession: B64372
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-276 <BU>
A:Cross-References: GB:U67506; GB:L77117; NID:91591274; PIDN:AAB98569.1; PID:91591285; T
C:Genetics:
A:Map position: REV514290-513460
A:Start codon: TTG
C:Superfamily: cell division inhibitor related protein; ferredoxin 2[4Fe-4S] homology
F:69-119/Domain: ferredoxin 2[4Fe-4S] homology <FER>

Query Match 100.0%; Score 28; DB 1; Length 276;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5
|||||
Db 212 FNVEX 216

RESULT 7
T11070
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - nematode (*Onchocerca volvulus*) m
C:Species: mitochondrion *Onchocerca volvulus*
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 12-Nov-1999
C:Accession: T11070
R:Kieddle, E.M.; Higazi, T.; Unasch, T.R.
Mol. Biochem. Parasitol. 95, 111-127, 1998
A:Title: The mitochondrial genome of *Onchocerca volvulus*: Sequence, structure and phylog
A:Reference number: Z17246; MUID:98434257
A:Accession: T11070
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-299 <KED>
A:Cross-References: EMBL:AF015193; NID:92735934; PID:92735942; PIDN:AAC61617.1
A:Experimental source: strain forest
C:Genetics:
A:Genome: mitochondrion
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 299;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5
|||||
Db 206 FNVEX 210

RESULT 8
G96944
probable permease [imported] - *Clostridium acetobutylicum*
C:Species: *Clostridium acetobutylicum*
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: G96944
R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clo*
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G96944
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <KUR>
A:Cross-References: GB:AE001437; PIDN:AAK78346.1; PID:915023214; GSPDB:GM00168
A:Experimental source: *Clostridium acetobutylicum* ATCC824
C:Genetics:
A:Gene: CAC0366

Query Match 100.0%; Score 28; DB 2; Length 301;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5
|||||
Db 51 FNVEX 55

RESULT 9
S58995
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - earthworm (*Lumbricus terrestr*
C:Species: mitochondrion *Lumbricus terrestris* (common earthworm)
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 07-Dec-1999
C:Accession: S58995
R:Boore, J.L.; Brown, W.M.
Genetics 141, 305-319, 1995
A:Title: Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus ter*
A:Reference number: S58985; MUID:96042914
A:Accession: S58995
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-308 <BOO>
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC4
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 308;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5
|||||
Db 211 FNVEX 215

RESULT 10
S01191
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - fruit fly (*Drosophila melanoga*
C:Species: mitochondrion *Drosophila melanogaster*
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 02-Nov-2001
C:Accession: S01191
R:Carlesse, R.
Genetics 118, 649-663, 1988
A:Title: *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionar
A:Reference number: S01185; MUID:88212147
A:Accession: S01191
A:Molecule type: DNA
A:Residues: 1-312 <GAR>
A:Note: the author translated the codon CTA for residue 190 as Ser and TCT for residu
C:Genetics:
A:Gene: ND-1
A:Cross-References: FlyBase:FBgn0013679
A:Genome: mitochondrion
A:Genetic code: SGC4
A:Start codon: ATA
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5
|||||

Db 215 FNVEY 219

RESULT 11

B71390

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - common lancelet mitochondrion
C:Species: mitochondrion Branchiostoma lanceolatum (common lancelet)

C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000

C:Accession: B71390

R:Spruyt, N.; Delarbre, C.; Gachejin, G.; Laudet, V.

Nucleic Acids Res. 26, 3279-3285, 1998

A:Title: Complete sequence of the amphioxus (Branchiostoma lanceolatum) mitochondrial ge
A:Reference number: A71390; MUID:98292550

A:Accession: B71390

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-313 <SPR>

A:Cross-references: GB:Y16474; NID:g3292989; PIDN:CAA76247.1; PID:g3292991

C:Genetics:

A:Gene: NADH1

A:Genome: mitochondrion

A:Genetic code: SGC4

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 313;
Best Local Similarity 100.0%; Pred. No. 36;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
|||||

Db 213 FNVEY 217

RESULT 12

T12010

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - Anopheles quadrimaculatus A mitoc
C:Species: mitochondrion Anopheles quadrimaculatus A

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 11-Jan-2000

C:Accession: T12010

R:Cockburn, A.F.; Mitchell, S.E.; Seawright, J.A.

Arch. Insect Biochem. Physiol. 14, 31-36, 1990

A:Title: Cloning of the mitochondrial genome of Anopheles quadrimaculatus.

A:Reference number: Z17375; MUID:92190510

A:Accession: T12010

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-314 <CCCC>

A:Cross-references: EMBL:104272; NID:g342501; PID:g1217663; PIDN:AAA93552.1

A:Experimental source: strain Orlando

C:Genetics:

A:Genome: mitochondrion

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 314;
Best Local Similarity 100.0%; Pred. No. 36;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
|||||

Db 216 FNVEY 220

RESULT 13

T09813

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - African malaria mosquito mitochondrion
C:Species: mitochondrion Anopheles gambiae (African malaria mosquito)

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Oct-2000

C:Accession: T09813

R:Beard, C.B.; Hamm, D.M.; Collins, F.H.

Insect Mol. Biol. 2, 103-124, 1993

A:Title: The mitochondrial genome of the mosquito Anopheles gambiae: DNA sequence, ge

A:Reference number: Z16863; MUID:97242550

A:Accession: T09813

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-314 <BEA>

A:Cross-references: EMBL:L20934; NID:g309056; PIDN:AAI2202.1; PID:g309069; GSPDB:GNO

A:Experimental source: strain G13

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SGC4

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 314;
Best Local Similarity 100.0%; Pred. No. 36;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
|||||

Db 216 FNVEY 220

RESULT 14

OXMSIM

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - mouse mitochondrion
N:Alternate names: NADH-ubiquinone oxidoreductase chain 1

C:Species: mitochondrion Mus musculus (house mouse)

C:Date: 02-Apr-1982 #sequence_revision 17-Jul-1998 #text_change 07-Dec-1999

C:Accession: A00409

R:Bibb, M.J.; Van Etten, R.A.; Wright, C.T.; Walberg, M.W.; Clayton, D.A.

Cell 26, 167-180, 1981

A:Title: Sequence and gene organization of mouse mitochondrial DNA.

A:Reference number: A00153; MUID:82137051

A:Accession: A00409

A:Molecule type: DNA

A:Residues: 1-315 <BBB>

A:Cross-references: GB:J01420; NID:g342520; PIDN:AA84644.1; PID:g896295

A:Note: the authors translated the initiation codon ATT for residue 1 as Ile

C:Genetics:

A:Gene: ND1

A:Genome: mitochondrion

A:Genetic code: SGC1

A:Start codon: ATT

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 1; Length 315;
Best Local Similarity 100.0%; Pred. No. 36;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
|||||

Db 208 FNVEY 212

RESULT 15

T11337

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - aardvark mitochondrion
C:Species: mitochondrion Oryzteropus afer (aardvark)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: T11337

R:Arnasson, U.; Gullberg, A.; Janke, A.

Proc. R. Soc. Lond. B Biol. Sci. 266, 339-345, 1999

A:Title: The mitochondrial DNA molecule of the aardvark, Oryzteropus afer, and the po

A:Reference number: Z17263; MUID:99197468

A:Accession: T11337

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-317 <ARN>

A:Cross-references: EMBL:Y18475; NID:g4691353; PIDN:CAB41621.1; PID:g4691354
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
A>Note: NADH1
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 FNVEY 5
11111
Db 211 FNVEY 215

Search completed: August 20, 2002, 11:10:34
Job time: 5664 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:32:59 ; Search time 55.29 Seconds

(without alignments)
3.501 Million cell updates/sec

Title: US-09-824-286-13

Perfect score: 28
Sequence: 1 FNVEY 5

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	175	1	RL6A_YEAST
2	28	100.0	175	1	RL6B_YEAST
3	28	100.0	259	1	CAH2_BOVIN
4	28	100.0	276	1	Y578_METJA
5	28	100.0	301	1	NYM1_NYCNO
6	28	100.0	303	1	NYM1_CABIN
7	28	100.0	303	1	NYM1_ZAGBR
8	28	100.0	308	1	NYM1_LUMTE
9	28	100.0	312	1	NYM1_DROME
10	28	100.0	312	1	NYM1_DROSU
11	28	100.0	313	1	NYM1_BRALA
12	28	100.0	314	1	NYM1_ANOGA
13	28	100.0	314	1	NYM1_ANOCU
14	28	100.0	315	1	NYM1_MOUSE
15	28	100.0	318	1	NYM1_BALMU
16	28	100.0	318	1	NYM1_BALPH
17	28	100.0	318	1	NYM1_BOVIN
18	28	100.0	318	1	NYM1_BRAVA
19	28	100.0	318	1	NYM1_CANFA
20	28	100.0	318	1	NYM1_CERST
21	28	100.0	318	1	NYM1_DASNO
22	28	100.0	318	1	NYM1_DIDMA
23	28	100.0	318	1	NYM1_EQUAS
24	28	100.0	318	1	NYM1_HALGR
25	28	100.0	318	1	NYM1_HIPAM
26	28	100.0	318	1	NYM1_HORSE
27	28	100.0	318	1	NYM1_LEMCA
28	28	100.0	318	1	NYM1_MACRO
29	28	100.0	318	1	NYM1_MACRO
30	28	100.0	318	1	NYM1_MACRO
31	28	100.0	318	1	NYM1_MANTE
32	28	100.0	318	1	NYM1_MONDO
33	28	100.0	318	1	NYM1_NOTTY

34	28	100.0	318	1	NYM1_NYCNO	078697 nycticebus
35	28	100.0	318	1	NYM1_ORNAN	037717 ornithorhyn
36	28	100.0	318	1	NYM1_PERGU	078710 peromyscus
37	28	100.0	318	1	NYM1_PRACT	078706 phascogaster
38	28	100.0	318	1	NYM1_PHOVI	000505 phoca vitul
39	28	100.0	318	1	NYM1_PIG	079874 sus scrofa
40	28	100.0	318	1	NYM1_RABIT	079427 oryctolagus
41	28	100.0	318	1	NYM1_RAF	003889 rattus norv
42	28	100.0	318	1	NYM1_RHIN	096189 rhinoceros
43	28	100.0	318	1	NYM1_SARHA	078711 sarcophilus
44	28	100.0	318	1	NYM1_SHEEP	078747 ovis aries
45	28	100.0	318	1	NYM1_SWIMA	078712 smilnopsids

ALIGNMENTS

RESULT 1
ID RL6A_YEAST STANDARD: PRT: 175 AA.
AC 002326;
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1993 (Rel. 37, Last annotation update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L6-A (L17) (YL16) (RP18).
OS RL6A OR YL16A OR YML073C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93003323; PubMed=1390890;
RA Hashimoto T., Suzuki K., Mizuta K., Otake E.;
RT "Yeast ribosomal proteins: XIV. Complete nucleotide sequences of the
two genes encoding Saccharomyces cerevisiae YL16.";
RL Biochim. Biophys. Acta 1132:195-198(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Brown D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L6 IN YEAST.
CC -1- SIMILARITY: BELONGS TO THE L6E FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D10225; BAA01077.1; -;
CC EMBL: Z46373; CAA86505.1; -;
CC PIR: S28944; S28944.
CC SGD: S0004538; RPL6A.
CC InterPro: IPR000915; Ribosomal_L6E.
CC Pfam: PF01159; Ribosomal_L6E; 1.
CC ProDom: PD009612; Ribosomal_L6E; 1.
CC PROSITE: PS01170; RIBOSOMAL_L6E; 1.
CC Ribosomal protein; Multigene family.
FT INIT MET 0
SQ SPROUCE 175 AA; 19830 MW; 0388034AAFT73BCB4 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 175;

Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
DB 100 FNVEY 104

```

RESULT 2
ID RLB6_YEAST STANDARD: PRT: 175 AA.
AC P05739:
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L6-B (L17) (YL16) (RP18).
GN RLB6 OR YL16B OR RPL16B OR YLR448W OR L9324.4.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93003323; PubMed=1390890;
RA Hashimoto T., Suzuki K., Mizuta K., Otake E.;
RT "Yeast ribosomal proteins: XIV. Complete nucleotide sequences of the
RL two genes encoding Saccharomyces cerevisiae YL16."
RL Biochim. Biophys. Acta 1132:195-198(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favetto A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Merdis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifkin L., Riles L., Taich A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-24.
RA Otake E., Higo K.-I., Itoh T.;
RT "Yeast ribosomal proteins: VIII. Isolation of two proteins and
RL sequence characterization of twenty-four proteins from cytoplasmic
RT ribosomes."
RL Mol. Gen. Genet. 195:544-546(1984).
CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L6 IN YEAST.
CC -1- SIMILARITY: BELONGS TO THE L6E FAMILY OF RIBOSOMAL PROTEINS.
CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10226; BAA01078.1; -
DR EMBL: U22382; AAB67529.1; -
DR PIR: S11257; S11257.
DR PIR: S28945; S28945.
DR SGD: S0004440; RPL6B.
DR InterPro: IPR000915; Ribosomal_L6E.
DR Pfam: PF01159; Ribosomal_L6e; 1.
DR ProDom: PD009612; Ribosomal_L6E; 1.
DR PROSITE: PS01170; RIBOSOMAL_L6E; 1.
KM Ribosomal protein; Multigene family.
FT INIT MET 0
FT CONFLICT 67 67 P -> L (IN REF. 1).
SQ SEQUENCE 175 AA; 19855 MW; 689AFA0813DC6651 CRC64;

```

Query Match 100.0%; Score 28; DB 1; Length 175;
 Best Local Similarity 100.0%; Pred. No. 7.1; Indels 0; Gaps 0;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
 DB 100 FNVEY 104

```

RESULT 3
ID CAH2_BOVIN STANDARD: PRT: 259 AA.
AC P00921:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbonic anhydrase II (EC 4.2.1.1) (Carbonate dehydratase II) (CA-II).
GN CA2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RC TISSUE=Erythrocyte;
RX MEDLINE=77065798; PubMed=826282;
RA Sclaky M., Limozin N., Filippi-Foveau D., Guillan J.M.,
RA Laurent-Tabuse G.;
RT "Primary structure of bovine erythrocyte carbonic anhydrase
RT Cl. II. Complete sequence."
RL Biochimie 58:1071-1082(1976).
RN [2]
RP REVISIONS.
RX MEDLINE=77242599; PubMed=19093;
RA Guillan J.M., Limozin N., Mallet B., di Costanzo J., Charrel M.;
RT "Genetic independence of two forms of carbonic anhydrase from bovine
RT erythrocytes."
RL Biochimie 59:293-302(1977).
CC -1- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.
CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- MISCELLANEOUS: ONE MINOR AND TWO MAJOR FORMS WERE ISOLATED
CC CHROMATOGRAPHICALLY.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE
CC FAMILY.
CC -1- DATABASE: NAME=worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/CA.html".
DR PIR: A01144; CRB02.
DR HSP: P00918; IBV3.
DR InterPro: IPR001148; Carb_anhydrase.
DR Pfam: PF00194; carb_anhydrase; 1.
DR ProDom: PD000865; Carb_anhydrase; 1.
DR PROSITE: PS00162; EUK_CO2_ANNHYDRASE; 1.
KM Lyase; Zinc; Acetylation.
FT MOD RES 1 1 ACETYLATION.
FT METAL 93 93 ZINC (CATALYTIC).
FT METAL 95 95 ZINC (CATALYTIC).
FT METAL 118 118 ZINC (CATALYTIC).
FT VARIANT 56 56 R -> Q (IN ONE OF THE MAJOR FORMS).
SQ SEQUENCE 259 AA; 28980 MW; 8488644617BF32D8 CRC64;

```

Query Match 100.0%; Score 28; DB 1; Length 259;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
 DB 65 FNVEY 69

RESULT 4
 ID Y578_METJA STANDARD: PRT: 276 AA.
 AC Q57998;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein M0578.

GN M0578. Methanococcus jannaschii.
OS Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.T.,
Overleak R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,
Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RA Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii. Science 273:1058-1073(1996).
RT Science 273:1058-1073(1996).
RL -1- SIMILARITY: BELONGS TO THE BACTERIAL TYPE FERREDOXIN FAMILY.
CC -1- SIMILARITY: TO M.JANNASCHII M0579.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67506; AAB98569.1; -.
DR HSSP: P00198; IFCA.
DR TIGR: M0578; -.
DR InterPro: IPR001450; 4Fe4S-ferredoxin.
DR InterPro: IPR000707; Para.
DR Pfam: PF00037; fer4; 2.
DR Pfam: PF00991; Para; 1.
DR PROSITE: PS00198; 4Fe4S-FERREDOXIN; 1.
KW Hypothetical protein; Iron-sulfur; 4Fe-4S; ATP-binding;
KM Complete proteome.
FT NP_BIND 15 22 ATP (POTENTIAL).
FT METAL 76 76 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 79 79 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 82 82 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 86 86 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 101 101 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 104 104 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 107 107 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 111 111 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 276 AA; 30794 MW; 0F900CE7F2E88669 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 276;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
DB 212 FNVEY 216

RESULT 5
NUM1_NYCNO STANDARD; PRT; 301 AA.
AC O80003;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
GN NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3) (Fragment).
OS Myctalus noctula (Noctule).
OG Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Vespertilionidae;
OC Myctalus.
OX NCBI_TaxID=51300;
RN [1]
RP SEQUENCE FROM N.A.
RA Petit E., Excoffier L., Mayer F.;
RT "No evidence of bottlenecks in the post-glacial recolonization of
RT Europe by the noctule bat (*Myctalus noctula*).";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBS databases.
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF065104; AAC28345.1; -.
DR EMBL: AF065108; AAC28349.1; -.
DR EMBL: AF065109; AAC28350.1; -.
DR InterPro: IPR001694; Resp-chain_NADH_DH1.
DR Pfam: PF00146; NADHdh; 1.
DR PROSITE: PS00667; COMPLEXI_ND1_1; 1.
DR PROSITE: PS00668; COMPLEXI_ND1_2; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
FT NON_TER 1 301
SQ SEQUENCE 301 AA; 33957 MW; B294975131BB0D09 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 301;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
DB 205 FNVEY 209

RESULT 6
NUM1_CABUN STANDARD; PRT; 303 AA.
AC O78700;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3) (Fragment).
GN MYND1 OR NDI.
OS Cabassous undulatus (Southern naked-tailed armadillo).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Edentata; Dasypodidae; Cabassous.
OX NCBI_TaxID=48852;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98404151; PubMed=9732458;
RA Cao Y., Janke A., Waddell P.J., Westernman M., Takenaka O., Murata S.,
Okada N., Paabo S., Hasegawa M.;
RT "Conflict among individual mitochondrial proteins in resolving the
RT phylogeny of eutherian orders.";
RL J. Mol. Evol. 47:307-322(1998).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AB011217; BAA32109.1; -
DR InterPro: IPR001694; Resp-chain_NADH_DHL.
DR Pfam: PF00146; NADHdh.1.
DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE: PS00667; COMPLEX1_ND1_2; 1.
KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 303 AA; 33839 MW; 692D072836585532 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
Db 199 FNVEY 203

RESULT 7
NM_2ZAGBR STANDARD; PRT; 303 AA.
AC 078713;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3) (Fragment).
GN MND1 OR NDI.
OS Zaglossus bruijnii (Long-beaked echidna).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Monotremata; Techylosidae; Zaglossus.
OX NCBI_TaxID=33543;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98404151; PubMed=9732458;
RA Cao Y., Janke A., Muddell P.J., Westernan M., Takenaka O., Murata S.,
RA Orada N., Paabo S., Hasegawa M.;
RT "Conflict among individual mitochondrial proteins in resolving the
RT phylogeny of eutherian orders";
RL J. Mol. Evol. 47:307-322(1998).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AB011230; BAA32122.1; -
DR InterPro: IPR001694; Resp-chain_NADH_DHL.
DR Pfam: PF00146; NADHdh.1.
DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE: PS00667; COMPLEX1_ND1_2; 1.
KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 303 AA; 34223 MW; 64E8AF098E26FC9F CRC64;

Query Match 100.0%; Score 28; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
Db 199 FNVEY 203

RESULT 8
NM_2LUMTE STANDARD; PRT; 308 AA.
AC 037546;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN NDI.
OS Lumbricus terrestris (Common earthworm).
OC Mitochondrion.
OC Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxidae;
OC Lumbricina; Lumbricidae; Lumbricus.
OX NCBI_TaxID=6398;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96042914; PubMed=8536978;
RA Boore J.L., Brown W.M.;
RT "Complete sequence of the mitochondrial DNA of the annelid worm
RT Lumbricus terrestris";
RL Genetics 141:305-319(1995).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U24570; AAC46874.1; -
DR InterPro: IPR001694; Resp-chain_NADH_DHL.
DR Pfam: PF00146; NADHdh.1.
DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE: PS00667; COMPLEX1_ND1_2; 1.
KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 308 AA; 34841 MW; 6FAE24B35D566DBC CRC64;

Query Match 100.0%; Score 28; DB 1; Length 308;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
Db 211 FNVEY 215

RESULT 9
NM_2DROME STANDARD; PRT; 312 AA.
AC P18929;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN MT:ND1 OR NDI.
OS Drosophila melanogaster (Fruit fly).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SQ SEQUENCE FROM N.A.
RC STRAIN=BRETAGNE;

RX MEDLINE=88212147; PubMed=3130291;
RA Garesse R.;
RT "Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations.";
RL Genetics 118:649-663(1988).
CC
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: M37275; AAA69715.1; -;
CC EMBL: U37341; AAC47823.1; -;
CC PIR: S01191; S01191.
CC FlyBase: FBgn0013679; mt:ND1.
CC InterPro: IPR001694; Resp_chain_NADH_DH1.
CC Pfam: PF00146; NADhdh.1
CC PROSITE: PS00667; COMPLEX1_ND1_1; 1.
CC PROSITE: PS00668; COMPLEX1_ND1_2; 1.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 312 AA; 35910 MW; B2B01B6213CE722 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FENVY 5
Db 215 FENVY 219

RESULT 10
NMJM_DROSU STANDARD; PRT; 312 AA.
AC P51937;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN MT:ND1 OR ND1.
OS Drosophila subobscura (Fruit fly).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7241;
RN [1]
RP SEQUENCE OF 1-152 FROM N.A.
RA MEDLINE=95106285; PubMed=7528808;
RA Barrio E., Latorre A., Moya A.;
RT "Phylogeny of the Drosophila obscura species group deduced from mitochondrial DNA sequences.";
RL J. Mol. Evol. 39:478-488(1994).
RN [2]
RP SEQUENCE OF 50-312 FROM N.A.
RC STRAIN=SSP. TUE 3;
RA Volz-Lingenhl A.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL: U07288; AAA76626.1; -;
CC DR EMBL: X65130; CAA46260.1; -;
CC DR FlyBase: FBgn0012956; Dsub\mt:ND1.
CC DR InterPro: IPR001694; Resp_chain_NADH_DH1.
CC DR Pfam: PF00146; NADhdh.1.
CC DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
CC DR PROSITE: PS00668; COMPLEX1_ND1_2; 1.
CC KM Mitochondrion; Oxidoreductase; NAD; Transmembrane.
SQ SEQUENCE 312 AA; 35877 MW; AFD1B3BC2BCD868 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FENVY 5
Db 215 FENVY 219

RESULT 11
NMJM_BRALA STANDARD; PRT; 313 AA.
AC Q21000; O47433;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN ND1 OR NAD1 OR NADH1.
OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus), and Branchiostoma floridae (Florida lancelet) (Amphioxus).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7740; 7739;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=B.lanceolatum;
RA MEDLINE=97398704; PubMed=9254918;
RA Delarbre C., Barriel V., Tillet S., Janvier P., Gachelin G.;
RT "The main features of the cranial mitochondrial DNA between the ND1 and the COI genes were established in the common ancestor with the lancelet.";
RL Mol. Biol. Evol. 14:807-813(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=B.lanceolatum;
RX MEDLINE=98292550; PubMed=9628930;
RA Spruyt N., Delarbre C., Gachelin G., Lauget V.;
RT "Complete sequence of the amphioxus (Branchiostoma lanceolatum) mitochondrial genome: relations to vertebrates.";
RL Nucleic Acids Res. 26:3279-3285(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=B.floridae;
RX MEDLINE=99261652; PubMed=10331267;
RA Boore J.L., Daehler L.L., Brown W.M.;
RT "Complete sequence, gene arrangement, and genetic code of mitochondrial DNA of the cephalochordate Branchiostoma floridae (Amphioxus).";
RL Mol. Biol. Evol. 16:410-418(1999).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL: Y09524; CAA70708.1; -
DR EMBL: Y16474; CAA76247.1; -
DR EMBL: AF098298; AAB88002.1; -
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADHdh: 1.
DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE: PS00667; COMPLEX1_ND1_2; 1.
DR OXidoreductase; NAD: Ubiquinone; Mitochondrion; Transmembrane.
FT CONFLICT 111 111 S->P (IN REF. 1).
FT CONFLICT 274 274 R->S (IN REF. 1).
SQ SEQUENCE 313 AA; 34429 MW; 1787EA3049F0B959 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 28; DB 1; Length 313;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNV5Y 5
Db 213 FNV5Y 217

```

```

RESULT 12
ID NUIK_ANOGA STANDARD; PRT; 314 AA.
AC P34846;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN NDI.
OS Anopheles gambiae (African malaria mosquito).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G3;
RX MEDLINE=97242550; PubMed=9087549;
RA Beard C.B., Hamm D.M., Collins F.H.;
RT "The mitochondrial genome of the mosquito Anopheles gambiae: DNA
RT sequence, genome organization, and comparisons with mitochondrial
RT sequences of other insects.";
RL Insect Mol. Biol. 2:103-104(1993).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L20934; AAD12202.1; -
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADHdh: 1.
DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE: PS00667; COMPLEX1_ND1_2; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 314 AA; 36100 MW; 512CABE0AA4E1BD7 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 28; DB 1; Length 314;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNV5Y 5

```

```

Db 216 FNV5Y 220

```

```

RESULT 13
ID NUIK_ANOOU STANDARD; PRT; 314 AA.
AC P33502;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN NDI.
OS Anopheles quadrimaculatus (Mosquito).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7166;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORLANDO;
RX MEDLINE=92190510; PubMed=2134168;
RA Cockburn A.F., Mitchell S.E., Seawright J.A.;
RT "Cloning of the mitochondrial genome of Anopheles quadrimaculatus.";
RL Arch. Insect Biochem. Physiol. 14:31-36(1990).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L04272; AAA93552.1; -
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADHdh: 1.
DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE: PS00667; COMPLEX1_ND1_2; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 314 AA; 36138 MW; 7574D2C4C78E4F35 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 28; DB 1; Length 314;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 FNV5Y 5
Db 216 FNV5Y 220

```

```

RESULT 14
ID NUIK_MOUSE STANDARD; PRT; 315 AA.
AC P03888;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN MTND1 OR MT-ND1 OR NDI.
OS Mus musculus (Mouse).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82137051; PubMed=7332926;
RA Bibb M.J., van Etten R.A., Wright C.T., Walberg M.W., Clayton D.A.;
RT "Sequence and gene organization of mouse mitochondrial DNA.";

```

RL Cell 26:167-180(1981).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: V00711; CAA24080.1; -;
 DR EMBL: J01420; AAB48644.1; -;
 DR PIR: A00409; OXMSIM.
 DR MGI: 101787; mt-ND1.
 DR InterPro: IPR001694; Resp_chain_NADH_DH1.
 DR Pfam: PF00146; NADHdh; 1.
 DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
 DR PROSITE: PS00668; COMPLEX1_ND1_2; 1.
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
 SQ SEQUENCE 315 AA; 35651 MW; B5018F305AEC3B3A CRC64;

Query Match 100.0%; Score 28; DB 1; Length 315;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FIVEY 5
 Db 208 FIVEY 212

RESULT 15
 NTJM_BALMU
 ID NTJM_BALMU STANDARD; PRT; 318 AA.
 AC P41296;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
 GN MTND1 OR ND1 OR NADH1.
 OS Balanoptera musculus (blue whale).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balanopteridae; Balanoptera.
 OX NCBI_TaxId=9771;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94141932; PubMed-8308901;
 RA Arnason U., Gulberg A.;
 RT "Comparison between the complete mtDNA sequences of the blue and the
 RT fin whale, two species that can hybridize in nature."
 RL J. Mol. Evol. 37:312-322(1993).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X72204; CAA50995.1; -;
 DR PIR: S41820; S41820.
 DR InterPro: IPR001694; Resp_chain_NADH_DH1.
 DR Pfam: PF00146; NADHdh; 1.
 DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
 DR PROSITE: PS00668; COMPLEX1_ND1_2; 1.
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.

SQ SEQUENCE 318 AA; 35765 MW; 8BFB423C219BA57E CRC64;
 Query Match 100.0%; Score 28; DB 1; Length 318;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FIVEY 5
 Db 211 FIVEY 215

Search completed: August 20, 2002, 11:33:00
 Job time: 1449 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:31:52 ; Search time 191.14 Seconds
(without alignments)
4.525 Million cell updates/sec

Title: US-09-824-286-13

Perfect score: 28
Sequence: 1 FNVEY 5

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertibrate:*
14: sp_unclassified:*
15: sp_rv1rus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	32	8 Q94VL5	Q94VL5 salmo trutt
2	28	100.0	32	8 Q94NY1	Q94NY1 salmo salar
3	28	100.0	80	10 Q9AW31	Q9AW31 guillardi
4	28	100.0	114	8 Q63938	Q63938 bufo gargar
5	28	100.0	114	8 Q63937	Q63937 bufo gargar
6	28	100.0	114	8 Q64007	Q64007 bufo andrew
7	28	100.0	114	8 Q63322	Q63322 bufo exsul
8	28	100.0	114	8 Q63324	Q63324 bufo viridi
9	28	100.0	114	8 Q63326	Q63326 bufo virid
10	28	100.0	114	8 Q63328	Q63328 bufo andrew
11	28	100.0	114	8 Q63330	Q63330 bufo gargar
12	28	100.0	118	8 Q9MG11	Q9MG11 coregonus a
13	28	100.0	119	8 Q9MDE1	Q9MDE1 oncorhynch
14	28	100.0	119	8 Q9MOW4	Q9MOW4 oncorhynch
15	28	100.0	121	8 Q958M0	Q958M0 rana catesb
16	28	100.0	121	8 Q958L7	Q958L7 rana sylvat

ALIGNMENTS

17	28	100.0	121	8 Q958L4	Q958L4 rana tempor
18	28	100.0	121	8 Q958L1	Q958L1 rana boylii
19	28	100.0	121	8 Q958K8	Q958K8 rana pretio
20	28	100.0	121	8 Q958K5	Q958K5 rana aurora
21	28	100.0	121	8 Q958K2	Q958K2 rana cascad
22	28	100.0	121	8 Q958J9	Q958J9 rana muscos
23	28	100.0	124	1 Q94PN0	Q94PN0 rana muscos
24	28	100.0	124	1 Q94PG3	Q94PG3 methanosarc
25	28	100.0	126	8 Q9BAK2	Q9BAK2 notophthalm
26	28	100.0	126	8 Q9BAW9	Q9BAW9 lytiocrito
27	28	100.0	126	8 Q9BAW6	Q9BAW6 pachytriton
28	28	100.0	126	8 Q9BAW3	Q9BAW3 triturus vu
29	28	100.0	126	8 Q9BAW0	Q9BAW0 chioglossa
30	28	100.0	126	8 Q9BAV7	Q9BAV7 merensieill
31	28	100.0	126	8 Q9BAV4	Q9BAV4 merensieill
32	28	100.0	126	8 Q9BAV2	Q9BAV2 salamandra
33	28	100.0	126	8 Q9BAV0	Q9BAV0 salamandra
34	28	100.0	126	8 Q9BAU7	Q9BAU7 merensieill
35	28	100.0	126	8 Q9BAU4	Q9BAU4 merensieill
36	28	100.0	126	8 Q9BAU1	Q9BAU1 merensieill
37	28	100.0	126	8 Q9BAT9	Q9BAT9 merensieill
38	28	100.0	126	8 Q9BAT7	Q9BAT7 merensieill
39	28	100.0	126	8 Q9BAT5	Q9BAT5 merensieill
40	28	100.0	126	8 Q9BAT3	Q9BAT3 merensieill
41	28	100.0	126	8 Q9BAT0	Q9BAT0 merensieill
42	28	100.0	126	8 Q9BAS7	Q9BAS7 merensieill
43	28	100.0	126	8 Q9BAS4	Q9BAS4 merensieill
44	28	100.0	126	8 Q9BAS2	Q9BAS2 merensieill
45	28	100.0	127	8 Q9TGA0	Q9TGA0 shlnisaurus

RESULT 1

Q94VL5 PRELIMINARY; PRT; 32 AA.
AC Q94VL5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 1 (FRAGMENT).
GN ND1.
OS Salmo trutta (Brown trout).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8032;
RN [1]
RP SEQUENCE FROM N.A.
RA Arnold A.L., Knox D., Verspoor E.;
RT "mtDNA sequence from brown trout from the River Dee, Scotland;
RT Position 4447 to 4546.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF405331; AAK97505.1; -;
KW Mitochondrion.
FT NON_TER 1
FT NON_TER 32
SQ SEQUENCE 32 AA; 3425 MW; 90180CC34D5F08AB CRC64;

Query Match 100.0%; Score 28; DB 8; Length 32;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
|||||
Db 16 FNVEY 20

RESULT 2
Q94NY1

```

ID 094NY1 PRELIMINARY; PRT; 32 AA.
AC 094NY1:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 1 (FRAGMENT).
GN NDI.
OS Salmo salar (Atlantic salmon).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V59, AND V62;
RA Consuegra S., Garcia de Leaniz C., Serdio A.L., Knox D., Straus L.G.,
RT Gonzalez-Morales M.R., Verspoor E.;
RT "Mitochondrial DNA variation in Paleolithic and modern Atlantic salmon
from the Iberian Glacial Refugium."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arnold A.L., Knox D., Verspoor E.;
RT "mtDNA sequence from Atlantic salmon from the Saint John River,
RT Canada: Position 4447 to 4546."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Arnold A.L., Knox D., Verspoor E.;
RT "mtDNA sequence from Atlantic salmon from the River Torne, Sweden:
RT Position 4447 to 4546."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF385752; AAK67347.1; -
DR EMBL; AF385753; AAK67348.1; -
DR EMBL; AF405329; AAK97503.1; -
DR EMBL; AF405330; AAK97504.1; -
KM Mitochondrion.
FT NON_TER 1 1
FT NON_TER 32 32
SQ SEQUENCE 32 AA; 3425 MW; 90180CC34D5F08AB CRC64;

Query Match 100.0%; Score 28; DB 8; Length 32;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
Db 16 FNVEY 20

RESULT 3
O9AW31 PRELIMINARY; PRT; 80 AA.
AC O9AW31:
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 9.6 KDA PROTEIN.
OS Gulliardia theta (Cryptomonas phi).
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Gulliardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20087226; PubMed=10618395;
RA Zanner S., Fraunholz M., Wastl J., Penny S.L., Beaton M.,
RA Cavalier-Smith T., Maier U., Douglas S.;
RT "Chloroplast protein and centrosomal genes, a tRNA intron, and odd
RT telomeres in an unusually compact eukaryotic genome, the cryptomonad
RT nucleomorph."
RL Proc. Natl. Acad. Sci. U.S.A. 97:200-205(2000).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=21223349; PubMed=11323671;
RA Douglas S., Zanner S., Fraunholz M., Beaton M., Penny S., Deng L.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.;
RT "The highly reduced genome of an enslaved algal nucleus."
RL Nature 410:1091-1096(2001).
DR EMBL; AJ010592; CAC27040.1; -
KW Hypothetical protein.
SQ SEQUENCE 80 AA; 9620 MW; 592C0AEB96A457DD CRC64;

Query Match 100.0%; Score 28; DB 10; Length 80;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
Db 23 FNVEY 27

RESULT 4
O63938 PRELIMINARY; PRT; 114 AA.
AC O63938:
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
GN NDI.
OS Bufo gargarizans.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
OX NCBI_TaxID=30331;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VARIOUS STRAINS.
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RA Tunleyev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal."
RL Mol. Phylogenet. Evol. 9:80-87(1998).
DR EMBL; AF004535; AAD05148.1; -
DR EMBL; AF004533; AAD05144.1; -
DR EMBL; AF004534; AAD05146.1; -
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADhdh; 1.
KM Mitochondrion.
FT NON_TER 1 1
FT NON_TER 114 114
SQ SEQUENCE 114 AA; 12892 MW; FB68884EC1B90E97 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
Db 8 FNVEY 12

RESULT 5
O63937 PRELIMINARY; PRT; 114 AA.
AC O63937:
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-OC7-2001 (TReMBLrel. 18, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
GN NDI.
OS Bufo gargarizans.

```

OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
OX NCBI_TaxID=30331;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RT Tunlijev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
from the eastern escarpment of the Tibetan Plateau: a case of
vicariance and dispersal.";
RL Mol. Phylogenet. Evol. 9:80-87(1998).
DR EMBL; AF004532; AAD05142.1; -;
DR EMBL; AF004531; AAD05140.1; -;
DR InterPro: IPR001694; Resp_chain_NADH.DH1.
DR Pfam; PF00146; NADhdh; 1.
KW Mitochondrion.
FT NON_TER
SQ SEQUENCE 114 AA; 12844 MW; EF3884EDAA20E8C CRC64;

QY Query Match 100.0%; Score 28; DB 8; Length 114;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 FNV5 5
8 FNV5 12

RESULT 6
ID 064007 PRELIMINARY; PRT; 114 AA.
AC 064007;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
GN NDI.
OS Bufo andrewsi.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
OX NCBI_TaxID=61428;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RT Tunlijev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
from the eastern escarpment of the Tibetan Plateau: a case of
vicariance and dispersal.";
RL Mol. Phylogenet. Evol. 9:80-87(1998).
DR EMBL; AF004528; AAD05134.1; -;
DR EMBL; AF004527; AAD05132.1; -;
DR InterPro: IPR001694; Resp_chain_NADH.DH1.
DR Pfam; PF00146; NADhdh; 1.
KW Mitochondrion.
FT NON_TER
SQ SEQUENCE 114 AA; 12906 MW; FB7F832360690E97 CRC64;

QY Query Match 100.0%; Score 28; DB 8; Length 114;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 FNV5 5
8 FNV5 12

RESULT 7

063322
ID 063322 PRELIMINARY; PRT; 114 AA.
AC 063322;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
GN NDI.
OS Bufo exul.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
OX NCBI_TaxID=30330;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RT Tunlijev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
from the eastern escarpment of the Tibetan Plateau: a case of
vicariance and dispersal.";
RL Mol. Phylogenet. Evol. 9:80-87(1998).
DR EMBL; AF004524; AAD05126.1; -;
DR InterPro: IPR001694; Resp_chain_NADH.DH1.
DR Pfam; PF00146; NADhdh; 1.
KW Mitochondrion.
FT NON_TER
SQ SEQUENCE 114 AA; 12886 MW; 7693866FEB1D824 CRC64;

QY Query Match 100.0%; Score 28; DB 8; Length 114;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 FNV5 5
8 FNV5 12

RESULT 8
ID 063324 PRELIMINARY; PRT; 114 AA.
AC 063324;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
GN NDI.
OS Bufo viridis.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
OX NCBI_TaxID=30338;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RT Tunlijev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
from the eastern escarpment of the Tibetan Plateau: a case of
vicariance and dispersal.";
RL Mol. Phylogenet. Evol. 9:80-87(1998).
DR EMBL; AF004525; AAD05128.1; -;
DR InterPro: IPR001694; Resp_chain_NADH.DH1.
DR Pfam; PF00146; NADhdh; 1.
KW Mitochondrion.
FT NON_TER
SQ SEQUENCE 114 AA; 12990 MW; B08B6D3AF7E00643 CRC64;

QY Query Match 100.0%; Score 28; DB 8; Length 114;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
 |||||
 Db 8 FNVEY 12

RESULT 9
 063326 PRELIMINARY; PRT; 114 AA.

ID 063326;
 AC 063326;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
 GN NDI.
 OS Bufo verrucosissimus.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
 OX NCBI_TaxID=61429;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98140324; PubMed=9479697;
 RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
 RT "Phylogenetic relationships of toads in the Bufo bufo species group
 from the eastern escarpment of the Tibetan Plateau: a case of
 vicariance and dispersal.";
 RT Mol. Phylogenet. Evol. 9:80-87(1998).
 RL EMBL: AF004526; AAD05130.1; -
 DR InterPro: IPR001694; Resp-chain_NADH_DH1.
 DR Pfam: PF00146; NADHdh; 1.
 KW Mitochondrion.
 FT NON_TER 1
 SQ SEQUENCE 114 AA; 12914 MW; A3799D377A6A27E8 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
 |||||
 Db 8 FNVEY 12

RESULT 10
 063328 PRELIMINARY; PRT; 114 AA.

ID 063328;
 AC 063328;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
 GN NDI.
 OS Bufo andrewsi.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
 OX NCBI_TaxID=61428;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98140324; PubMed=9479697;
 RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
 RT "Phylogenetic relationships of toads in the Bufo bufo species group
 from the eastern escarpment of the Tibetan Plateau: a case of
 vicariance and dispersal.";
 RT Mol. Phylogenet. Evol. 9:80-87(1998).
 RL EMBL: AF004529; AAD05136.1; -
 DR InterPro: IPR001694; Resp-chain_NADH_DH1.
 DR Pfam: PF00146; NADHdh; 1.
 KW Mitochondrion.

FT NON_TER 1
 SQ SEQUENCE 114 AA; 12876 MW; E70F9F4F10690E82 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
 |||||
 Db 8 FNVEY 12

RESULT 11
 063330 PRELIMINARY; PRT; 114 AA.

ID 063330;
 AC 063330;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
 GN NDI.
 OS Bufo gargarizans.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
 OX NCBI_TaxID=30331;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98140324; PubMed=9479697;
 RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
 RT "Phylogenetic relationships of toads in the Bufo bufo species group
 from the eastern escarpment of the Tibetan Plateau: a case of
 vicariance and dispersal.";
 RT Mol. Phylogenet. Evol. 9:80-87(1998).
 RL EMBL: AF004530; AAD05138.1; -
 DR InterPro: IPR001694; Resp-chain_NADH_DH1.
 DR Pfam: PF00146; NADHdh; 1.
 KW Mitochondrion.
 FT NON_TER 1
 SQ SEQUENCE 114 AA; 12918 MW; E3CB884EC1B90E97 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5
 |||||
 Db 8 FNVEY 12

RESULT 12
 09MG11 PRELIMINARY; PRT; 118 AA.

ID 09MG11;
 AC 09MG11;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE NADH DEHYDROGENASE (FRAGMENT).
 GN NDI.
 OS Coregonus artedii.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Coregonus.
 OX NCBI_TaxID=36181;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Turgeon J., Bernatchez L.;
 RT "Coregonus artedii phylogeography.";
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF246933; AAF90053.1; -;
DR InterPro: IPR001694; Resp_chain_NADH_DH1.
DR Pfam: PF00146; NADHdh. 1.
DR PROSITE: PS00668; COMPLEX1_ND1_2; 1.
KW Mitochondrion.
FT NON_TER 1 1
FT NON_TER 118 118
SQ SEQUENCE 118 AA: 12880 MW: 71851F8DA57D4C09 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 118;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
|||||
DB 62 FNVEY 66

RESULT 13
O9MDE1 PRELIMINARY; PRT; 119 AA.
AC O9MDE1:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 1 (FRAGMENT).
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RA McCusker M.R., Parkinson E., Taylor E.B.;
RT "Mitochondrial DNA variation in rainbow trout (Oncorhynchus mykiss)
RT across its native range: testing biogeographical hypotheses and their
RT relevance to conservation."
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF254849; AAF71365.1; -;
DR EMBL: AF254843; AAF71359.1; -;
DR EMBL: AF254844; AAF71360.1; -;
DR EMBL: AF254845; AAF71361.1; -;
DR EMBL: AF254846; AAF71362.1; -;
DR EMBL: AF254847; AAF71363.1; -;
DR EMBL: AF254848; AAF71364.1; -;
DR InterPro: IPR001694; Resp_chain_NADH_DH1.
DR Pfam: PF00146; NADHdh. 1.
KW Mitochondrion.
FT NON_TER 1 1
SQ SEQUENCE 119 AA: 13174 MW: DED639949FE02174 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 119;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
|||||
DB 12 FNVEY 16

RESULT 14
O9MJM4 PRELIMINARY; PRT; 119 AA.
AC O9MJM4:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 1 (FRAGMENT).
OS Oncorhynchus clarki (Cutthroat trout) (Salmo clarki).
OG Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=30962;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20571467; PubMed=11123621;
RA McCusker M.R., Parkinson E., Taylor E.B.;
RT "Mitochondrial DNA variation in rainbow trout (Oncorhynchus mykiss)
RT across its native range: testing biogeographical hypotheses and their
RT relevance to conservation."
RL MOL. ECOL. 9:2089-2108(2000).
DR EMBL: AF254865; AAF71366.1; -;
DR InterPro: IPR001694; Resp_chain_NADH_DH1.
DR Pfam: PF00146; NADHdh. 1.
KW Mitochondrion.
FT NON_TER 1 1
SQ SEQUENCE 119 AA: 13193 MW: DF6789949EE12074 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 119;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
|||||
DB 12 FNVEY 16

RESULT 15
O958M0 PRELIMINARY; PRT; 121 AA.
AC O958M0:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
GN ND1.
OS Rana catesbeiana (Bull frog).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranae.
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21184280; PubMed=11286498;
RA Macey J.R., Strasburg J.L., Brissson J.A., Vredenburg V.T.,
RA Jennings M., Larson A.;
RT "Molecular Phylogenetics of Western North American Frogs of the Rana
RT boylii Species Group."
RT MOL. Phylogenet. Evol. 19:131-143(2001).
DR EMBL: AF314016; AAK56866.1; -;
KW Mitochondrion.
FT NON_TER 1 1
SQ SEQUENCE 121 AA: 13569 MW: D657A0BF338819FF CRC64;

Query Match 100.0%; Score 28; DB 8; Length 121;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
|||||
DB 16 FNVEY 20

Search completed: August 20, 2002, 11:31:53
Job time: 1467 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:07:15 ; Search time 187.18 Seconds
(without alignments)
4.154 Million cell updates/sec

Title: US-09-824-286-14
Perfect score: 38
Sequence: 1 KEIHLYO 7

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq_032802.*
2: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1980.DAT:*
3: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1981.DAT:*
4: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1982.DAT:*
5: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1983.DAT:*
6: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1984.DAT:*
7: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1985.DAT:*
8: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	7	19 AAW31651	Human cytokine rec
2	38	100.0	230	15 AAR47151	IL-2 receptor gamma
3	38	100.0	230	16 AAR82934	Interleukin 4 comp
4	38	100.0	232	15 AAR47150	IL-2 receptor gamma
5	38	100.0	347	15 AAR47149	IL-2 receptor gamma
6	38	100.0	369	15 AAR47148	IL-2 receptor gamma
7	38	100.0	482	19 AAW31646	Human cytokine rec
8	38	100.0	691	21 AAY92202	Fusion polypeptide
9	38	100.0	694	21 AAY92201	Fusion polypeptide
10	38	100.0	694	21 AAY92203	Fusion polypeptide
11	33	86.8	1059	21 AAB37570	Wheat starch synth

12	33	86.8	1628	21 AAB37569	Wheat starch synth
13	33	86.8	1628	21 AAB49304	Wheat starch synth
14	32	84.2	233	20 AAY07011	Breast cancer asso
15	32	84.2	233	21 AAB42565	Human OREF ORP2329
16	32	84.2	367	20 AAY31735	Human cell cycle r
17	31	81.6	71	21 AAG40993	zeae may's protein f
18	31	81.6	85	21 AAG18933	zeae may's protein f
19	31	81.6	120	21 AAG18932	zeae may's protein f
20	31	81.6	175	21 AAG34058	zeae may's protein f
21	31	81.6	210	21 AAG34057	zeae may's protein f
22	31	81.6	258	21 AAG34056	zeae may's protein f
23	31	81.6	263	22 AAG23964	Novel human diagno
24	31	81.6	677	22 AAB61419	Drosophila melanog
25	31	81.6	756	22 ABB58408	Drosophila melanog
26	31	81.6	785	22 ABB62933	Drosophila melanog
27	30	78.9	231	18 AAW20162	H. pylori inner me
28	30	78.9	431	21 AAB52596	Helicobacter pylor
29	30	78.9	433	17 AAW05196	Helicobacter pylor
30	30	78.9	438	18 AAW20916	H. pylori inner me
31	30	78.9	454	22 AAG92495	C glutamicum prote
32	29	76.3	38	22 AAM96286	Human reproductive
33	29	76.3	38	22 AAM42315	Human breast or ov
34	29	76.3	60	22 AAG99784	ERA binding domain
35	29	76.3	70	22 AAM18654	Peptide #508 enco
36	29	76.3	81	22 AAG40659	Propionibacterium
37	29	76.3	104	21 AAG57029	Arabidopsis thalia
38	29	76.3	107	21 AAG57028	Arabidopsis thalia
39	29	76.3	108	21 AAG57027	Arabidopsis thalia
40	29	76.3	111	22 AAG82931	S. epidermidis ope
41	29	76.3	111	22 AAG83036	S. epidermidis ope
42	29	76.3	133	22 AAM39731	Human polypeptide
43	29	76.3	161	22 AAM41517	Human polypeptide
44	29	76.3	194	21 AAG55688	Arabidopsis thalia
45	29	76.3	197	21 AAG55687	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AAW31651	AAW31651 standard; Peptide; 7 AA.
XX	XX
AC	AAW31651;
XX	XX
DT	21-MAY-1998 (first entry)
XX	XX
DE	Human cytokine receptor gc chain epitope.
XX	XX
KW	Cytokine receptor; gamma common chain; gc chain; human;
KW	blocking agent; monoclonal antibody; CP-88; immunological disease;
KW	myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
KW	insulin-dependent diabetes; inflammatory bowel disease;
KW	sympathetic ophthalmia; uveitis; allergy; asthma; infection;
KW	graft versus host disease; psoriasis; immunosuppressive; therapy;
epitope.	
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO9743416-A1.
XX	XX
PD	20-NOV-1997.
XX	XX
PF	09-MAY-1997; 97WO-US07870.
XX	XX
PR	10-MAY-1996; 96US-0017466.
XX	XX
PA	(BIOJ) BIOGEN INC.
XX	XX
PI	Benjamin CD, Burkly LC, Hession C, Whitty A;
XX	XX
DR	WPI: 1998-008885/01.
XX	XX

PT Blocking agents of the gamma common chain of cytokine receptors -
 PT particularly monoclonal antibodies, used to induce T cell anergy for
 PT treatment of immunological diseases
 XX
 PS Claim 24; Page 84; 11pp; English.
 XX
 CC This peptide comprises an epitope of the human cytokine receptor
 CC common gamma (gc) chain (see AAW31646) that is recognised by
 CC gc blocking agents of the invention. 5 Such epitopes (see
 CC AAW31650-54) have been identified. The invention provides
 CC compositions and methods for inhibiting cytokine signalling using
 CC gc chain blocking agents for the treatment of immunological
 CC diseases such as myasthenia gravis, rheumatoid arthritis, lupus,
 CC multiple sclerosis, insulin-dependent diabetes, inflammatory bowel
 CC disease, sympathetic ophthalmia, uveitis, allergy, asthma,
 CC parasitic infection, graft vs. host disease or psoriasis. A
 CC preferred gc blocking agent is Mab CP.88 or its Fab fragment (see
 CC also AAW31647-48).
 XX
 SQ Sequence 7 AA:
 Query Match 100.0%; Score 38; DB 19; Length 7;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KEIHLQY 7
 DQ 1 keihlyq 7
 RESULT 2
 AAR47151
 ID AAR47151 standard; Protein: 230 AA.
 AC AAR47151:
 XX
 XX 13-JUN-1994 (first entry)
 DT
 XX
 DE IL-2 receptor gamma chain.
 XX
 KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
 KW rheumatoid arthritis; transplant rejection; primer;
 KW polymerase chain reaction; PCR; amplification.
 XX
 OS Homo sapiens.
 XX
 XX EP578932-A.
 PN
 PD 19-JAN-1994.
 XX
 XX 22-APR-1993; 93EP-0106561.
 PF
 XX 23-APR-1992; 92JP-0104947.
 PR
 XX
 PA (AJIN) AJINOMOTO KK.
 PA (SUGA/) SUGAMURA K.
 XX
 PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
 PI Suzuki M, Takeshita T;
 XX
 XX WPI: 1994-017546/03.
 DR N-PSDB; AAQ54831.
 XX
 XX DNA and protein sequences of IL-2 gamma chain - useful as immune
 PT regulatory agents for treatment of e.g. rheumatoid arthritis and
 PT transplant rejection
 XX
 PS Disclosure; Page 22-23, 35-36, 50pp; English.
 XX
 CC The human IL-2 receptor gamma chain preform (AAR47148), including the
 CC signal peptide, is encoded by the sequence given in AAQ54828. The
 CC mature protein (AAR47149) is encoded by sequence AAQ54829. A soluble

CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AAQ54830.
 CC while a soluble form suitable for expression in prokaryotes (AAR47151)
 CC is encoded by AAQ54831. Primers 1-6 (AAQ54820-25) are based on the N-
 CC terminal sequence of IL-2 receptor gamma chain, and are used to
 CC isolate IL2 receptor gamma chain receptor cDNA. Primers AAQ54826-27
 CC are used to obtain the protein given in AAR47151.
 XX
 SQ Sequence 230 AA:
 Query Match 100.0%; Score 38; DB 15; Length 230;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KEIHLQY 7
 DQ 98 keihlyq 104
 RESULT 3
 AAR82934
 ID AAR82934 standard; Protein: 230 AA.
 AC AAR82934:
 XX
 XX 26-FEB-1996 (first entry)
 DT
 XX
 DE Interleukin 4 component common to the IL-2 receptor gamma chain.
 XX
 KW Interleukin-4; IL-4; gamma chain component; immunosuppressants;
 KW anti-allergy agent; signal transmisson inhibitor; autoimmune;
 KW disease; anti-inflammatory; anaphylactic shock; bronchial asthma;
 KW Interleukin-2; IL-2; atopic dermatitis; urticaria.
 XX
 OS Homo sapiens.
 XX
 XX JP07149662-A.
 PN
 XX
 PD 13-JUN-1995.
 XX
 XX 07-SEP-1994; 94JP-0213706.
 PF
 XX 08-SEP-1993; 93JP-0223574.
 PR
 XX
 PA (AJIN) AJINOMOTO KK.
 PA (SUGA/) SUGAMURA K.
 XX
 DR WPI: 1995-243601/32.
 DR N-PSDB; AAT04952.
 XX
 XX Novel interleukin-4 receptor monoclonal antibodies inhibit signal
 PT transmission - useful as immunosuppressants and anti-allergy agents.
 PT
 XX
 PS Example 1; Page 9; 11pp; Japanese.
 XX
 CC AAT04952 encodes AAR82934 a component of the IL-4 receptor common to
 CC the IL-2 receptor gamma chain molecule, which was used to generate
 CC anti-IL-4 receptor monoclonal antibodies (mabs). The mabs (IL-4
 CC signal transmisson inhibitors) can be used as immunosuppressants
 CC and anti-allergy agents, for the treatment of autoimmune and chronic
 CC inflammatory diseases, e.g. anaphylactic shock, bronchial asthma,
 CC atopic dermatitis and urticaria.
 XX
 SQ Sequence 230 AA:
 Query Match 100.0%; Score 38; DB 16; Length 230;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KEIHLQY 7
 DQ 99 keihlyq 105


```
RESULT 4
AA047150
ID AAR47150 standard; Protein: 252 AA.
XX
AC AAR47150;
XX
DT 13-JUN-1994 (first entry)
XX
DE IL-2 receptor gamma chain.
XX
KM Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KW rheumatoid arthritis; transplant rejection; primer;
XX polymerase chain reaction; PCR; amplification; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= Sig_Peptide
XX
PN EP578932-A.
XX
PD 19-JAN-1994.
XX
PE 22-APR-1993; 93EP-0106561.
XX
PR 23-APR-1992; 92JP-0104947.
XX
PA (AJIN ) AJINOMOTO KK.
XX (SUGA/) SUGAMURA K.
XX
PI Aseo H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
DR WPI: 1994-017546/03.
DR N-PSDB; AA054830.
XX
PT DNA and protein sequences of IL-2 gamma chain - useful as immune
PT regulatory agents for treatment of e.g. rheumatoid arthritis and
PT transplant rejection
XX
PS Disclosure; Page 21-22, 34-35; 50pp; English.
XX
CC The human IL-2 receptor gamma chain preform (AAR47148), including the
CC signal peptide, is encoded by the sequence given in AA054828. The
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
CC while a soluble form suitable for expression in prokaryotes (AAR47151)
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
CC terminal sequence of IL-2 receptor gamma chain, and are used to
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
CC are used to obtain the protein given in AAR47151.
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 38; DB 15; Length 252;
Best Local Similarly 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
Db 120 keihlyq 126

RESULT 5
AAR47149
ID AAR47149 standard; Protein: 347 AA.
XX
AC AAR47149;
XX
```

```
DT 13-JUN-1994 (first entry)
XX
DE IL-2 receptor gamma chain.
XX
KM Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KW rheumatoid arthritis; transplant rejection; primer;
XX polymerase chain reaction; PCR; amplification.
XX
OS Homo sapiens.
XX
PN EP578932-A.
XX
PD 19-JAN-1994.
XX
PE 22-APR-1993; 93EP-0106561.
XX
PR 23-APR-1992; 92JP-0104947.
XX
PA (AJIN ) AJINOMOTO KK.
XX (SUGA/) SUGAMURA K.
XX
PI Aseo H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
DR WPI: 1994-017546/03.
DR N-PSDB; AA054829.
XX
PT DNA and protein sequences of IL-2 gamma chain - useful as immune
PT regulatory agents for treatment of e.g. rheumatoid arthritis and
PT transplant rejection
XX
PS Claim 4; Page 41; 50pp; English.
XX
CC The human IL-2 receptor gamma chain preform (AAR47148), including the
CC signal peptide, is encoded by the sequence given in AA054828. The
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
CC while a soluble form suitable for expression in prokaryotes (AAR47151)
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
CC terminal sequence of IL-2 receptor gamma chain, and are used to
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
CC are used to obtain the protein given in AAR47151.
XX
SQ Sequence 347 AA;

Query Match 100.0%; Score 38; DB 15; Length 347;
Best Local Similarly 100.0%; Pred. No. 7.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
Db 98 keihlyq 104

RESULT 6
AAR47148
ID AAR47148 standard; Protein: 369 AA.
XX
AC AAR47148;
XX
DT 13-JUN-1994 (first entry)
XX
DE IL-2 receptor gamma chain.
XX
KM Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KW rheumatoid arthritis; transplant rejection; primer; PCR;
XX polymerase chain reaction; amplification; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
```

```

FT      /label= Sig-peptide
XX      EP578932-A.
XX      19-JAN-1994.
XX      22-APR-1993; 93EP-0106561.
XX      23-APR-1992; 92JP-0104947.
XX      (AJIN ) AJINOMOTO KK.
XX      (SUGA/) SUGAMURA K.
XX      Aseo H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
XX      Suzuki M, Takeshita T;
XX      WPI: 1994-017546/03.
XX      N-PSDB: AAQ54828.
XX      DNA and protein sequences of IL-2 gamma chain - useful as immune
XX      regulatory agents for treatment of e.g. rheumatoid arthritis and
XX      transplant rejection
XX      Disclosure; Page 16-17, 29-30; 50pp; English.
XX      The human IL-2 receptor gamma chain preform (AAR47148), including the
XX      signal peptide, is encoded by the sequence given in AAQ54828. The
XX      mature protein (AAR47149) is encoded by sequence AAQ54829. A soluble
XX      form of IL-2 receptor gamma chain (AAR47150) is encoded by AAQ54830,
XX      while a soluble form suitable for expression in prokaryotes (AAR47151)
XX      is encoded by AAQ54831. Primers 1-6 (AAQ54820-25) are based on the N-
XX      terminal sequence of IL-2 receptor gamma chain, and are used to
XX      isolate IL2 receptor gamma chain receptor cDNA. Primers AAQ54826-27
XX      are used to obtain the protein given in AAR47151.
XX      Sequence 369 AA:
SQ      Query Match 100.0%; Score 38; DB 15; Length 369;
        Best Local Similarity 100.0%; Pred. No. 8.5;
        Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 KEIHLQY 7
        |||||
        120 keihlyq 126
        Db
RESULT 7
ID      AAW31646 standard; Protein: 482 AA.
XX      AAW31646;
XX      21-MAY-1998 (first entry)
XX      Human cytokine receptor gc chain-Ig fusion protein.
XX      Cytokine receptor; gamma common chain; gc chain; human;
XX      blocking agent; monoclonal antibody; CP.B8; Immunological disease;
XX      myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
XX      insulin-dependent diabetes; inflammatory bowel disease;
XX      sympathetic ophthalmia; uveitis; allergy; asthma; infection;
XX      graft versus host disease; psoriasis; immunosuppressive; therapy.
XX      Chimeric - Homo sapiens.
XX      Key Location/Qualifiers
XX      FT 1..254
XX      FT /note= "gc chain N-terminal region"
XX      FT Protein 235..482
XX      FT /note= "IgG1 constant region"
XX      FT 255..264
XX      FT Region /note= "IgG1 hinge region"

```

```

FT      Domain 264..482
FT      /note= "IgG1 CH2 and CH3 constant domains1"
XX      W09743416-A1.
XX      20-NOV-1997.
XX      09-MAY-1997; 97WO-US07870.
XX      10-MAY-1996; 96US-0017466.
XX      (BIOJ ) BIOGEN INC.
XX      Benjamin CD, Burkly LC, Hession C, Whitty A;
XX      WPI: 1998-008885/01.
XX      N-PSDB: AAT97439.
XX      Blocking agents of the gamma common chain of cytokine receptors -
XX      particularly monoclonal antibodies, used to induce T cell anergy for
XX      treatment of immunological diseases
XX      Example 1; Page 79-80; 111pp; English.
XX      This polypeptide comprises a fusion between the N-terminal 254
XX      amino acids of the human mature cytokine receptor gamma common (gc)
XX      chain and the hinge region and CH2 and CH3 constant domains of
XX      human IgG1. The fusion was expressed from clone pLB001 (see
XX      AAY97439) in COS-7 cells, and used to generate murine anti-human gc
XX      specific monoclonal antibodies (MAbs), including CP.B8 produced by
XX      hybridoma ATCC HB 12107. The invention provides compositions and
XX      methods for inhibiting cytokine signalling using gc chain blocking
XX      agents for the treatment of immunological diseases such as
XX      myasthenia gravis, rheumatoid arthritis, lupus, multiple sclerosis,
XX      insulin-dependent diabetes, inflammatory bowel disease, sympathetic
XX      ophthalmia, uveitis, allergy, asthma, parasitic infection, graft
XX      vs. host disease or psoriasis. A preferred gc blocking agent is
XX      Mab CP.B8 or its Fab fragment (see also AAW31647-48).
XX      Sequence 482 AA:
SQ      Query Match 100.0%; Score 38; DB 19; Length 482;
        Best Local Similarity 100.0%; Pred. No. 11;
        Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 KEIHLQY 7
        |||||
        120 keihlyq 126
        Db
RESULT 8
ID      AAY92202 standard; Protein: 691 AA.
XX      AAY92202;
XX      01-AUG-2000 (first entry)
XX      Fusion polypeptide 603, IL-4 trap.
XX      IL-4 trap; cytokine; antagonist; CNRF; receptor; fusion protein;
XX      cyostatic; immunomodulator; osteopathic.
XX      Synthetic.
XX      OS Homo sapiens.
XX      Key Location/Qualifiers
XX      FT 1..254
XX      FT W0200018932-A2.
XX      FT 06-APR-2000.
XX      FT 22-SEP-1999; 99WO-US22045.

```

```

PR 25-SEP-1998; 980S-0101858.
PR 19-MAY-1999; 990S-0313942.
XX
PA (REGE-) REGENERON PHARM INC.
XX
PI Stahl N, Yancopoulos GD;
XX
DR WPI; 2000-293165/25.
DR N-PSDB; AAA09044.
XX
PT Isolated nucleic acid molecule for treating cytokine-related diseases
PT or disorders encodes a fusion polypeptide capable of binding a cytokine
PT to form a nonfunctional complex
XX
PS Example 6; Fig 22A-D; 152pp; English.
XX
CC This sequence shows fusion polypeptide 603, which is capable of
CC binding cytokine IL-4 to form a non-functional complex.
CC The invention concerns production of antagonists to any cytokine that
CC utilizes an alpha specificity determining component, which when combined
CC with the cytokine, binds to a first beta signal transducing component to
CC form a non-functional intermediate which then binds to a second beta
CC signal transducing component causing beta-receptor dimerization, the
CC soluble alpha specificity determining component of the receptor
CC (SR-alpha) and the extracellular domain of the first beta signal
CC transducing component of the cytokine receptor (beta-1) are combined to
CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the
CC cytokine by binding the cytokine to form a non-functional complex. The
CC receptor components are shared by cytokines such as the CNTF (ciliary
CC neurotrophic factor) family of cytokines. The invention provides the
CC basis for the development of IL-6 antagonists, as they show that if, in
CC the presence of a ligand, a non-functional intermediate complex,
CC consisting of the ligand, its alpha receptor and its beta-1 receptor
CC component, can be formed, it will effectively block the action of the
CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers
CC of the extracellular domains of the alpha specificity determining
CC components of their receptors and the extracellular domain of gp130.
CC The resultant heterodimers, function as high-affinity traps, rendering
CC the cytokine inaccessible to form a signal transducing complex with the
CC native membrane-bound forms of their receptor. The nucleic acids and
CC polypeptides are useful for treating cytokine-related diseases or
CC disorders such as osteoporosis and primary and secondary effects of
CC cancer including multiple myeloma or cachexia.
XX
SQ Sequence 691 AA;

Query Match 100.0%; Score 38; DB 21; Length 691;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLQ 7
Db 120 keihlyg 126

RESULT 9
AA92201
ID AAY92201 standard; Protein; 694 AA.
XX
AC AAY92201;
XX
DT 01-AUG-2000 (first entry)
XX
DE Fusion polypeptide 424, IL-4 trap.
XX
KW IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;
KW cytostatic; immunomodulator; osteopathic.
XX
OS Synthetic.
XX Homo sapiens.
XX
PN WO200018932-A2.

```

```

XX
XX 06-APR-2000.
PD
XX
XX 22-SEP-1999; 99WO-US22045.
PE
XX
XX 25-SEP-1998; 980S-0101858.
PR 19-MAY-1999; 990S-0313942.
XX
PA (REGE-) REGENERON PHARM INC.
XX
PI Stahl N, Yancopoulos GD;
XX
DR WPI; 2000-293165/25.
DR N-PSDB; AAA09043.
XX
PT Isolated nucleic acid molecule for treating cytokine-related diseases
PT or disorders encodes a fusion polypeptide capable of binding a cytokine
PT to form a nonfunctional complex
XX
PS Example 6; Fig 21; 152pp; English.
XX
CC This sequence shows fusion polypeptide 424, which is capable of
CC binding cytokine IL-4 to form a non-functional complex.
CC The invention concerns production of antagonists to any cytokine that
CC utilizes an alpha specificity determining component, which when combined
CC with the cytokine, binds to a first beta signal transducing component to
CC form a non-functional intermediate which then binds to a second beta
CC signal transducing component causing beta-receptor dimerization, the
CC soluble alpha specificity determining component of the receptor
CC (SR-alpha) and the extracellular domain of the first beta signal
CC transducing component of the cytokine receptor (beta-1) are combined to
CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the
CC cytokine by binding the cytokine to form a non-functional complex. The
CC receptor components are shared by cytokines such as the CNTF (ciliary
CC neurotrophic factor) family of cytokines. The invention provides the
CC basis for the development of IL-6 antagonists, as they show that if, in
CC the presence of a ligand, a non-functional intermediate complex,
CC consisting of the ligand, its alpha receptor and its beta-1 receptor
CC component, can be formed, it will effectively block the action of the
CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers
CC of the extracellular domains of the alpha specificity determining
CC components of their receptors and the extracellular domain of gp130.
CC The resultant heterodimers, function as high-affinity traps, rendering
CC the cytokine inaccessible to form a signal transducing complex with the
CC native membrane-bound forms of their receptor. The nucleic acids and
CC polypeptides are useful for treating cytokine-related diseases or
CC disorders such as osteoporosis and primary and secondary effects of
CC cancer including multiple myeloma or cachexia.
XX
SQ Sequence 694 AA;

Query Match 100.0%; Score 38; DB 21; Length 694;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLQ 7
Db 120 keihlyg 126

RESULT 10
AA92203
ID AAY92203 standard; Protein; 694 AA.
XX
AC AAY92203;
XX
DT 01-AUG-2000 (first entry)
XX
DE Fusion polypeptide 622, IL-4 trap.
XX
KW IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;
KW cytostatic; immunomodulator; osteopathic.
XX

```

XX OS Synthetic.
 XX OS Homo sapiens.
 XX PN WO200018932-A2.
 XX PD 06-APR-2000.
 XX PF 22-SEP-1999; 99WO-US22045.
 XX PR 25-SEP-1998; 98US-0101858.
 XX PR 19-MAY-1999; 99US-0313942.
 XX PA (REGG-) REGENERON PHARM INC.
 XX PI Stahl N, Yancopoulos GD;
 XX DR WPI: 2000-293165/25.
 XX DR N-PSDB; AAA09045.
 XX PS Example 6; Fig 23A-D; 152pp; English.
 CC This sequence shows fusion polypeptide 622, which is capable of
 CC binding cytokine IL-4 to form a non-functional complex.
 CC The invention concerns production of antagonists to any cytokine that
 CC utilizes an alpha specificity determining component, which when combined
 CC with the cytokine, binds to a first beta signal transducing component to
 CC form a non-functional intermediate which then binds to a second beta
 CC signal transducing component causing beta-receptor dimerization, the
 CC soluble alpha specificity determining component of the receptor
 CC (SR-alpha) and the extracellular domain of the first beta signal
 CC transducing component of the cytokine receptor (beta-1) are combined to
 CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the
 CC cytokine by binding the cytokine to form a non-functional complex. The
 CC receptor components are shared by cytokines such as the CNTF (ciliary
 CC neurotrophic factor) family of cytokines. The invention provides the
 CC basis for the development of IL-6 antagonists, as they show that if, in
 CC the presence of a ligand, a non-functional intermediate complex,
 CC consisting of the ligand, its alpha receptor and its beta-1 receptor
 CC component, can be formed, it will effectively block the action of the
 CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers
 CC of the extracellular domains of the alpha specificity determining
 CC components of their receptors and the extracellular domain of gp130.
 CC The resultant heterodimers, function as high-affinity traps, rendering
 CC the cytokine inaccessible to form a signal transducing complex with the
 CC native membrane-bound forms of their receptor. The nucleic acids and
 CC polypeptides are useful for treating cytokine-related diseases or
 CC disorders such as osteoporosis and primary and secondary effects of
 CC cancer including multiple myeloma or cachexia.
 XX SQ Sequence 694 AA;
 OY 1 KEIHLQ 7
 Db 120 keihlyq 126
 Query Match 100.0%; Score 38; DB 21; Length 694;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX DE wheat starch synthase III SEQ ID NO: 10.
 XX KW wheat; starch synthase; SSII; SSIII; starch content; starch synthetase;
 XX KW food product; adhesive.
 XX OS Trilicium aestivum.
 XX PN WO200066745-A1.
 XX PD 09-NOV-2000.
 XX PF 28-APR-2000; 2000WO-AU00385.
 XX PR 29-APR-1999; 99AU-0000052.
 XX PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 XX PA (GOOD-) GOODMAN FIELDER LTD.
 XX PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.
 XX PI Morell M, Li Z, Rahman S, Appels R;
 XX DR WPI: 2000-647602/62.
 XX DR N-PSDB; AAC86414.
 XX PS Nucleic acid molecules encoding wheat starch synthase (WST)-I and
 XX WST-II, useful in modifying plant starch content and/or composition -
 XX Claim 9; Page 183-186; 211pp; English.
 CC The present invention relates to novel protein and coding sequences from
 CC wheat. The proteins are wheat starch synthases, designated SSII and
 CC SSIII. These can be used in the modification of plant starch content or
 CC composition, and to screen plants to identify mutations which affect
 CC starch content and composition. The starch can then be used in food
 CC products, such as flour, and in films, coatings, adhesives, building
 CC materials and packaging materials.
 XX SQ Sequence 1059 AA;
 OY 1 KEIHLQ 7
 Db 663 kdihlyq 669
 Query Match 86.8%; Score 33; DB 21; Length 1059;
 Best Local Similarity 71.4%; Pred. No. 2,8e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
 AAB37569
 ID AAB37569 standard; Protein; 1628 AA.
 XX AC AAB37569;
 XX DT 01-MAR-2001 (first entry)
 XX DE wheat starch synthase III SEQ ID NO: 8.
 XX DE wheat; starch synthase; SSII; SSIII; starch content; starch synthetase;
 XX KW food product; adhesive.
 XX OS Trilicium aestivum.
 XX PN WO200066745-A1.
 XX PD 09-NOV-2000.
 XX PF 28-APR-2000; 2000WO-AU00385.
 XX PR 29-APR-1999; 99AU-0000052.

PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PA (GOOD-) GOODMAN FIELDER LTD.
 PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.
 XX
 PI Morell M, Li Z, Rahman S, Appels R;
 XX
 DR WPI; 2000-647602/62.
 DR N-PSDB; AAC86413.
 PT Nucleic acid molecules encoding wheat starch synthase (WST)-I and
 PT WST-II, useful in modifying plant starch content and/or composition -
 XX
 PS Claim 9; Page 172-179; 211pp; English.
 XX
 CC The present invention relates to novel protein and coding sequences from
 CC wheat. The proteins are wheat starch synthases, designated SSII and
 CC SSIII. These can be used in the modification of plant starch content or
 CC composition, and to screen plants to identify mutations which affect
 CC starch content and composition. The starch can then be used in food
 CC products, such as flour, and in films, coatings, adhesives, building
 CC materials and packaging materials.
 XX
 SQ Sequence 1628 AA;

Query Match 86.8%; Score 33; DB 21; Length 1628;
 Best Local Similarity 71.4%; Pred. No. 4.4e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLVQ 7
 I::IIII
 Db 1232 kdhllyg 1238

RESULT 13

AAB49304 standard; Protein; 1628 AA.

AC AAB49304;

DT 01-MAR-2001 (first entry)

DE Wheat starch synthase III protein.

KW Wheat; starch synthase; SSII; SSIII; starch content; starch synthesis;

KW food product; adhesive.

OS Triticum aestivum.

PN WO200066745-A1.

PD 09-NOV-2000.

PF 28-APR-2000; 2000WO-AU00385.

PR 29-APR-1999; 99AU-0000052.

PA (CSIR) COMMONWEALTH SCI & IND RES ORG.

PA (GOOD-) GOODMAN FIELDER LTD.

PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.

PI Morell M, Li Z, Rahman S, Appels R;

DR WPI; 2000-647602/62.

Nucleic acid molecules encoding wheat starch synthase (WST)-I and
 WST-II, useful in modifying plant starch content and/or composition -
 Example 13; Fig 7; 211pp; English.

The present invention relates to novel protein and coding sequences from
 wheat. The proteins are wheat starch synthases, designated SSII and
 SSIII. These can be used in the modification of plant starch content or

composition, and to screen plants to identify mutations which affect
 starch content and composition. The starch can then be used in food
 products, such as flour, and in films, coatings, adhesives, building
 materials and packaging materials.

SQ Sequence 1628 AA;

Query Match 86.8%; Score 33; DB 21; Length 1628;
 Best Local Similarity 71.4%; Pred. No. 4.4e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLVQ 7
 I::IIII
 Db 1232 kdhllyg 1238

RESULT 14

AAV07011 standard; Protein; 164 AA.

AC AAV07011;

DT 02-JUL-1999 (first entry)

DE Breast cancer associated antigen precursor sequence.

KW Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

KW prostate cancer.

OS Homo sapiens.

PN WO9904265-A2.

PD 28-JAN-1999.

PF 15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

PA (LUDM-) LUDMIG INST CANCER RES.

PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;

PI Tureci O;

DR WPI; 1999-132448/11.

PF New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers

PS Disclosure; Page 393-394; 787pp; English.

The invention relates to a method for diagnosing a disorder characterised
 by expression of a human cancer associated antigen precursor coded for by
 a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 biological sample isolated from a subject with an agent that specifically
 binds to the NAM, an expression product or a fragment of an expression
 product complexed with an HLA molecule; and (b) determining the
 interaction between the agent and the NAM or the expression product as a
 determination of the disorder. The products and methods can be used in
 the diagnosis, monitoring, research, or treatment of conditions
 characterised by the expression of various cancer associated antigens.
 The invention provides nucleic acid sequences and encoded polypeptides
 which are cancer associated antigen precursors expressed in human breast
 cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

CC lung cancer.
SQ Sequence 164 AA;
Query Match 84.2% Score 32; DB 20; Length 164;
Best Local Similarity 71.4% Pred. NO. 58;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEIHLQ 7
:|||||
Db 99 eeihlyq 105
RESULT 15
AAB42565
ID AAB42565 standard; Protein: 233 AA.
XX
AC AAB42565;
XX
DF 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2329 polypeptide sequence SEQ ID NO:4658.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineoplastic;
KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antineoplastic disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PE 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0340763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
DR WPI: 2000-602362/57.
DR N-PSDB: AAC6774.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 3855-3856; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antineoplastic; antibacterial; antiviral; antifungal; antineumatic;

CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 233 AA;
XX
Query Match 84.2% Score 32; DB 21; Length 233;
Best Local Similarity 71.4% Pred. NO. 85;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEIHLQ 7
:|||||
Db 193 eeihlyq 199
Search completed: August 20, 2002, 11:07:16
Job time: 6101 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:10:34 ; Search time 100.84 Seconds
(without alignments)
6.670 Million cell updates/sec

Title: US-09-824-286-14
Perfect score: 38
Sequence: 1 KEIHLYQ 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	369	2 A42565	interleukin-2 rece
2	32	84.2	257	2 J00403	hypothetical 28.1k
3	32	84.2	312	4 A40473	RNA-directed RNA p
4	32	84.2	433	2 A10536	probable membrane
5	32	84.2	551	2 T05667	probable growth re
6	31	81.6	233	2 S75267	hypothetical prote
7	31	81.6	294	2 H82272	conserved hypothet
8	31	81.6	373	2 A55718	interleukin-2 rece
9	31	81.6	472	2 S28026	nuclear pore compl
10	31	81.6	594	2 T18628	hypothetical prote
11	31	81.6	673	2 H69828	ABC transporter (A
12	30	78.9	135	2 B95172	hypothetical prote
13	30	78.9	135	2 B98038	hypothetical prote
14	30	78.9	188	2 S74963	hypothetical prote
15	30	78.9	213	2 E83420	endopeptidase C1p
16	30	78.9	215	1 VCMGPM	coat protein - pap
17	30	78.9	223	2 F86728	beta-subunit of L-
18	30	78.9	330	2 D90482	dehydrogenase, pro
19	30	78.9	339	2 AB3267	hypothetical prote
20	30	78.9	357	2 S21209	4-hydroxyphenylpyr
21	30	78.9	433	2 G64594	hemolysin secretio
22	30	78.9	433	2 H71917	methylin-accepting c
23	30	78.9	456	2 T40416	hypothetical prote
24	30	78.9	940	2 H82159	probable helicase
25	30	78.9	1322	2 H86196	hypothetical prote
26	30	78.9	4196	2 T43274	dynein heavy chain
27	29	76.3	129	2 B82745	hypothetical prote
28	29	76.3	146	2 T15533	hypothetical prote
29	29	76.3	153	2 D64628	hypothetical prote

30	29	76.3	153	2 E71888	hypothetical prote
31	29	76.3	194	2 C96740	hypothetical prote
32	29	76.3	197	2 T33661	hypothetical prote
33	29	76.3	259	2 AC1110	conserved hypothet
34	29	76.3	259	2 AG1471	conserved hypothet
35	29	76.3	328	2 S39599	class I histocompa
36	29	76.3	347	2 T32768	hypothetical prote
37	29	76.3	357	2 T32881	hypothetical prote
38	29	76.3	380	2 JC6321	CMP-NeuN-CGM3 alp
39	29	76.3	380	2 A56950	alpha-2,8-sialytra
40	29	76.3	387	2 S77268	carboxynorspermidl
41	29	76.3	425	2 E84631	probable serine ca
42	29	76.3	484	2 A12527	hypothetical prote
43	29	76.3	491	2 G83850	ATP-dependent DNA
44	29	76.3	548	2 C82839	translation releas
45	29	76.3	713	2 H64464	hypothetical prote

ALIGNMENTS

RESULT 1

A42565
interleukin-2 receptor gamma chain - human
C:Species: Homo sapiens (man)
C:Date: 04-Mar-1993 #sequence.revision 18-Nov-1994 #text_change 20-Jun-2000
C:Accession: A42565; A46591; I54332
R:Takeshita, T.; Asao, H.; Ohtani, K.; Ishii, N.; Kumaki, S.; Tanaka, N.; Munakata, I
Science 257, 379-382, 1992
A:Title: Cloning of the gamma chain of the human IL-2 receptor.
A:Reference number: A42565; WUID:92335883
A:Accession: A42565
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid; protein
A:Residues: 1-369 <TRK>
A:Cross-references: GB:011086; NID:q303611; PIDN:BA01857.1; PID:q219690
A:Experimental source: MOLR beta lymphoid cells
A:Note: sequence extracted from NCBI backbone (NCBIP:109167)
R:Noguchi, M.; Adelstein, S.; Cao, X.; Leonard, W.J.
J. Biol. Chem. 268, 13601-13608, 1993
A:Title: Characterization of the human interleukin-2 receptor gamma chain gene.
A:Reference number: A46591; WUID:93293887
A:Accession: A46591
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369 <RRS>
A:Cross-references: GB:011283; NID:q307056; PIDN:AA59145.1; PID:q307058
R:Puck, J.M.; Deschenes, S.M.; Porter, J.C.; Dutra, A.S.; Brown, C.J.; Willard, H.F.,
Hum. Mol. Genet. 2, 1099-1104, 1993
A:Title: The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-1
A:Reference number: I54332; WUID:94004847
A:Accession: I54332
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369 <RB2>
A:Cross-references: GB:0119546; NID:q349631; PIDN:AAC37524.1; PID:q349632
C:Genetics:
A:Gene: GDB:112RG; SCIDX1; IMD4
A:Cross-references: GDB:134807; OMIM:308380
A:Map position: Xq13.1-Xq13.1
A:Intons: 39/1; 90/2; 152/1; 198/3; 253/1; 285/2; 308/3
A:Note: defects are associated with an X-linked form of severe combined immunodeficit
C:Superfamily: interleukin-2 receptor gamma chain
C:Keywords: cytokine receptor; duplication; immunodeficiency; severe combined immunon

Query Match 100.0%; Score 38; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLYQ 7
DB 120 KEIHLYQ 126

RESULT 2
J00403
hypothetical 28.1k protein - clover yellow mosaic virus
N:Contains: coat protein
C:Species: clover yellow mosaic virus
C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 20-Jun-2000
R:Accession: J00403
R:Adonaidar, M.G.; Lai, R.
J. Gen. Virol. 70, 1871-1875, 1989
A:Title: Nucleotide sequence of the 3'-terminal region of clover yellow mosaic virus RNA
A:Reference number: J00402; MUID:89293092
A:Accession: J00403
A:Molecule type: genomic RNA
A:Residues: 1-257 <ABO>
A:Cross-references: GB:D00465; NID:g221218; PIDN:BAA00373.1; PID:g221220
A>Note: It is possible that the coat protein of CYMV is first translated as a 28.1k prot
C:Superfamily: potato virus coat protein
F:46-257/Product: coat protein #status predicted <COP>

Query Match 84.2%; Score 32; DB 2; Length 257;
Best Local Similarity 71.4%; Pred. No. 27;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
|:||||
Db 216 KNVHLXQ 222

RESULT 3
A40473
RNA-directed RNA polymerase (EC 2.7.7.48) / coat protein mutant fusion protein - clover
N:Contains: coat protein; RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: clover yellow mosaic virus
C>Date: 29-Jan-1993 #sequence_revision 24-Apr-1997 #text_change 19-May-2000
R:Accession: A40473
R:White, K.A.; Bancroft, J.B.; Mackie, G.A.
Virology 183, 479-486, 1991
A:Title: Defective RNAs of clover yellow mosaic virus encode nonstructural/coat protein
A:Reference number: A40473; MUID:91306431
A:Accession: A40473
A:Molecule type: genomic RNA
A:Residues: 1-312 <WHI>
A:Cross-references: GB:M63511; NID:g323436; PIDN:AAA42935.1; PID:g323437
A>Note: fragments of three other mutant fusion proteins are presented
C:Keywords: coat protein; fusion protein; nucleotidyltransferase
F:1-221/Region: RNA-directed RNA polymerase
F:222-312/Region: coat protein

Query Match 84.2%; Score 32; DB 4; Length 312;
Best Local Similarity 71.4%; Pred. No. 33;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
|:||||
Db 271 KNVHLXQ 277

RESULT 4
A10536
Probable membrane protein STY0305 [imported] - Salmonella enterica subsp. enterica serov
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
R:Accession: A10536
R:Perkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: A10536
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08737.1; PID:916501559; GSPDB:GN00176
C:Genetics:
A:Gene: STY0305

Query Match 84.2%; Score 32; DB 2; Length 433;
Best Local Similarity 71.4%; Pred. No. 46;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
|:||||
Db 96 REMHLXQ 102

RESULT 5
T05667
probable growth regulator F22113.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
R:Accession: T05667
R:Bevan, M.; Medler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer,
submitted to the Protein Sequence Database, February 1999
A:Reference number: T15420
A:Accession: T05667
A:Molecule type: DNA
A:Residues: 1-551 <BEV>
A:Cross-references: EMBL:AL035539
A:Experimental source: cultivar Columbia; BAC clone F22113
C:Genetics:
A:Map position: 4
A:Introns: 85/3; 116/1; 141/3; 171/2; 234/3; 295/2; 329/3; 440/2
A>Note: F22113.160

Query Match 84.2%; Score 32; DB 2; Length 551;
Best Local Similarity 83.3%; Pred. No. 59;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLX 6
|:||||
Db 360 KEVHLX 365

RESULT 6
S75267
hypothetical protein slr1039 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
R:Accession: S75267
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O.; Kaneo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S75267
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <KAN>
A:Cross-references: EMBL:D90904; GB:AB001339; NID:91652225; PIDN:BAA17181.1; PID:d101
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: bioc homology
F:43-148/Domain: bioc homology <BIOC>

Query Match 81.6%; Score 31; DB 2; Length 233;

Best Local Similarity 83.3%; Pred. No. 39;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 EHLHYQ 7
1:|||||
DB 208 EHLHYQ 213

RESULT 7
H82272
conserved hypothetical protein VC0853 [imported] - Vibrio cholerae (strain N16961 serogr
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: H82272
R/Heidelberg, J.F.; Elsen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.
L., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A82035; MUID:20406833
A/Accession: H82272
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-294 <HEI>
A/Cross-references: GB:AE004170; GB:AE003852; NID:g9655298; PIDN:AAF94015.1; GSPDB:GN001
A/Experimental source: serogroup O1; strain N16961; biotype El Tor
C/Genetics:
A/Gene: VC0853
A/Map position: 1
C/Superfamily: conserved hypothetical protein HI0072

Query Match 81.6%; Score 31; DB 2; Length 294;
Best Local Similarity 83.3%; Pred. No. 50;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 EHLHYQ 7
1:|||||
DB 259 EHLHYQ 264

RESULT 8
A55718
Interleukin-2 receptor gamma chain precursor - dog
C/Species: Canis lupus familiaris (dog)
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 23-Jul-1999
C/Accession: A55718
R/Henthorn, P.S.; Somberg, R.L.; Fintiani, V.M.; Puck, J.M.; Patterson, D.F.; Felsburg, F.
Genomics 23, 69-74, 1994
A/Title: IL-2Rgamma gene microdeletion demonstrates that canine X-linked severe combined
A/Reference number: A55718; MUID:95130114
A/Accession: A55718
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-373 <HEN>
A/Cross-references: GB:U04361; NID:9517411; PIDN:AAC48403.1; PID:9517412
C/Superfamily: interleukin-2 receptor gamma chain
C/Keywords: cytokine receptor; duplication

Query Match 81.6%; Score 31; DB 2; Length 373;
Best Local Similarity 71.4%; Pred. No. 64;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEIHLHYQ 7
1:|||||
DB 120 KEIHLHYQ 126

RESULT 9
S28026
nuclear pore complex protein NUP49 - yeast (Saccharomyces cerevisiae)
N/Alternate names: nuclear pore complex protein NSP49; nucleoporin NSP49; protein G1648;

C/Species: Saccharomyces cerevisiae
C/Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000
C/Accession: S28026; C44402; S52458; S59239; S64189; S28538
R/Wimmer, C.; Doye, V.; Grandi, P.; Nehrbass, U.; Hurt, E.C.
EMBO J. 11, 5051-5061, 1992
A/Title: A new subclass of nucleoporins that functionally interact with nuclear pore
A/Reference number: S28026; MUID:93098880
A/Accession: S28026
A/Molecule type: DNA
A/Residues: 1-472 <NIM>

A/Cross-references: EMBL:X68109; NID:g4055; PIDN:CAA8229.1; PID:g4056
R/Wente, S.R.; Rout, M.P.; Blobel, G.
J. Cell Biol. 119, 705-723, 1992
A/Title: A new family of yeast nuclear pore complex proteins.
A/Reference number: A44402; MUID:93054906
A/Accession: C44402

A/Molecule type: DNA
A/Residues: 1-472 <MEN>
A/Cross-references: EMBL:Z15040; NID:g4077; PIDN:CAV78758.1; PID:g4078
A/Note: sequence extracted from NCBI backbone (NCBIP:117140)
R/Bertani, I.; Coglievina, M.; Zaccaria, P.; Klima, R.; Bruschi, C.V.
Submitted to the EMBL data library, February 1995
A/Description: The sequence of 11.1kb fragment on the left arm to Saccharomyces cere
A/Reference number: S52454
A/Accession: S52458
A/Molecule type: DNA
A/Residues: 1-472 <BER>

A/Cross-references: EMBL:X84705; NID:g677853; PID:g677857
R/Bertani, I.; Coglievina, M.; Zaccaria, P.; Klima, R.; Bruschi, C.V.
Yeast 11, 1187-1194, 1995
A/Title: The sequence of an 11.1 kb fragment on the left arm of Saccharomyces cerevi
A/Reference number: S59235; MUID:96109931
A/Accession: S59239
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-472 <BER>

A/Cross-references: EMBL:X84705; NID:g677853; PIDN:CAA59181.1; PID:g677857
R/Bruschi, C.V.; Coglievina, M.; Bertani, I.; Klima, R.; Zaccaria, P.; Delneri, D.
Submitted to the Protein Sequence Database, May 1996
A/Reference number: S64183
A/Accession: S64189
A/Molecule type: DNA
A/Residues: 1-472 <BRU>

A/Cross-references: EMBL:Z72694; NID:g1322775; PID:g1322776; MIPS:YGL172w
A/Experimental source: strain S288C
C/Genetics:
A/Gene: SGD:NUP49; NSP49
A/Cross-references: SGD:S0003140; MIPS:YGL172w
A/Map position: 7L
C/Keywords: nucleus

Query Match 81.6%; Score 31; DB 2; Length 472;
Best Local Similarity 71.4%; Pred. No. 82;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEIHLHYQ 7
1:|||||
DB 375 KEIHLHYQ 381

RESULT 10
T18628
hypothetical protein B0001.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T18628
R/Sims, M.
Submitted to the EMBL data library, February 1996
A/Reference number: Z18999
A/Accession: T18628
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA

A:Residues: 1-594 <NLU>
A:Cross-references: EMBL:269634; PIDN:CAA93452.1; GSPDB:GN00022; CESP:B0001.3
A:Experimental source: clone B0001
A:Genetics:
A:Gene: CESP:B0001.3
A:Map position: 4
A:Introns: 9/2; 20/3; 69/2; 109/3; 154/3; 226/2; 340/3; 493/3; 555/3

Query Match
Best Local Similarity 81.6%; Score 31; DB 2; Length 594;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
||:||||
D5 217 KDVLHYE 223

RESULT 11

H69828
ABC transporter (ATP-binding protein) homolog yheH - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001

C:Accession: H69828

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Berte
C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.: Ehlich, S.D.; Emmerston, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fuma, S.; Galizzi, A.; Gall
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Kocher, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kuita, K.; Lapidis, A.; Ladiouis,
A.: Authors: Lauber, J.; Lazarevic, V.; Lee, S.K.; Levine, A.; Liu, H.; Masuda, S.; Meue
Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schreier, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: H69828
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-673 <KUN>
A:Cross-references: GB:299109; GB:AL009126; NID:92633260; PIDN:CAB12811.1; PID:ell82973;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yheH
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: ATP; nucleotide binding; P-loop
F:446-640/Domain: ATP-binding cassette homology <ABC>
F:463-470/Region: nucleotide-binding motif A (P-loop)

Query Match
Best Local Similarity 81.6%; Score 31; DB 2; Length 673;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
||:||||
D5 86 KEAHIYQ 92

RESULT 12

B95172

hypothetical protein SPI477 [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: B95172

R:Letellin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Hickey, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95172
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-135 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75571.1; PID:q14972967; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI477

Query Match
Best Local Similarity 78.9%; Score 30; DB 2; Length 135;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
||:||||
D5 89 KELYLYQ 95

RESULT 13

B98038

hypothetical protein spr1331 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: B98038

R:Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAlren, S.
J.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: B98038

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-135 <KUR>

A:Cross-references: GB:AE007317; PIDN:JAL00135.1; PID:q15458976; GSPDB:GN00174

C:Genetics:
A:Gene: spr1331

Query Match
Best Local Similarity 78.9%; Score 30; DB 2; Length 135;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
||:||||
D5 89 KELYLYQ 95

RESULT 14

S74963

hypothetical protein slr1599 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S74963

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O, K.; Okumura, S.; Shingo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
S.

A:Reference number: S74322; MUID:97061201

A:Accession: S74963

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-188 <KAN>

A:Cross-references: EMBL:DB0902; GB:AB001339; NID:q1652027; PIDN:BAI17003.1; PID:d101
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:

A:Start codon: GTG

Query Match

78.9%; Score 30; DB 2; Length 188;

Best Local Similarity 83.3%; Pred. No. 52;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 EIHLYQ 7

|||||

DB 64 EIHLYE 69

RESULT 15

E83420

endopeptidase Clp (EC 3.4.21.92) chain P PA1801 [similarity] - Pseudomonas aeruginosa (S

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-sep-2000 #text_change 03-Aug-2001

C:Accession: E83420

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lofy, S.; Olson, M.V.

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: AB2950; M01D:20437337

A:Accession: E83420

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-213 <STO>

A:Cross-references: GB:AE004606; GB:AE004091; NID:99947780; PIDN:AA05190.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: clpP; PA1801

C:Superfamily: endopeptidase Clp chain P

C:Keywords: hydrolase; serine proteinase

F:114/Active site: Ser #status predicted

F:139/Active site: His #status predicted

Query Match

78.9%; Score 30; DB 2; Length 213;

Best Local Similarity 83.3%; Pred. No. 59;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLY 6

|||||

DB 74 KDTHLY 79

Search completed: August 20, 2002, 11:10:36
Job time: 5666 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:33:00 ; Search time 55.29 seconds

(Without alignments)
4.902 Million cell updates/sec

Title: US-09-824-286-14

Perfect score: 38

Sequence: 1 KEIHLXQ 7

Scoring table: BIOSUM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	369	1	CYRG_HUMAN
2	32	84.2	212	1	COAT_CYMV
3	31	81.6	51	1	YORO_TTV1
4	31	81.6	212	1	PSB2_ORISA
5	31	81.6	294	1	PNK_VIBCH
6	31	81.6	373	1	CYRG_CANFA
7	31	81.6	472	1	MU49_YEAST
8	31	81.6	756	1	CPSB_DROME
9	30	78.9	50	1	REPA_BPT4
10	30	78.9	213	1	CLP1_PSEAE
11	30	78.9	215	1	COAT_PMV
12	30	78.9	357	1	HPPD_PSESP
13	30	78.9	4196	1	DYHC_SCHPO
14	29	76.3	223	1	YGB5_PASMU
15	29	76.3	286	1	YGB5_EDMIC
16	29	76.3	380	1	CAGD_MOUSE
17	29	76.3	411	1	RAPS_CHICK
18	29	76.3	882	1	YB90_MYCTU
19	29	76.3	3674	1	SPCR_HUMAN
20	28	73.7	152	1	NDKB_MOUSE
21	28	73.7	197	1	COAE_BACNO
22	28	73.7	260	1	FAPR_ECOLI
23	28	73.7	359	1	PST_CRIGR
24	28	73.7	359	1	PST_HUMAN
25	28	73.7	359	1	PST_MOUSE
26	28	73.7	377	1	PRGR_SHEEP
27	28	73.7	415	1	CGA2_XENLA
28	28	73.7	431	1	T232_BACTB
29	28	73.7	433	1	APL5_HUMAN
30	28	73.7	581	1	SYP_CHLTR
31	28	73.7	628	1	V70K_TYMC
32	28	73.7	628	1	V70K_TYMC
33	28	73.7	656	1	YC26_PORPU

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	369 AA.
CYRG_HUMAN				
ID	CYRG_HUMAN			
AC	P31785:			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cytokine receptor common gamma chain precursor (gamma-C) (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64) (CD132 antigen).			
GN	IL2RG			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92355883; PubMed=1631559.			
RA	Takeshita T., Asao H., Ohtani K., Ishii N., Kumaki S., Tanaka N.,			
RA	Munakata H., Nakamura M., Sugamura K.;			
RT	"Cloning of the gamma chain of the human IL-2 receptor.";			
RL	Science 257:379-382(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=93293887; PubMed=8514792;			
RA	Noguchi M., Adelstein S., Cao X., Leonard W.J.;			
RT	"Characterization of the human Interleukin-2 receptor gamma chain			
RT	gene.";			
RL	J. Biol. Chem. 268:13601-13608(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND VARIANTS ASP-114 AND ASN-153.			
RX	MEDLINE=94004847; PubMed=8401490;			
RA	Puck J.M., Deschenes S.M., Porter J.C., Dutra A.S., Brown C.J.,			
RA	Willard H., Henthorn P.S.;			
RT	"The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated			
RT	in X-linked severe combined immunodeficiency, SCIDX1.";			
RL	Hum. Mol. Genet. 2:1099-1104(1993).			
RN	[4]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RX	MEDLINE=9400315; PubMed=826076;			
RA	Kondo M., Takeshita T., Ishii N., Nakamura M., Watanabe S.,			
RA	Arai K., I., Sugamura K.;			
RT	"Sharing of the Interleukin-2 (IL-2) receptor gamma chain between			
RT	receptors for IL-2 and IL-4.";			
RL	Science 262:1874-1877(1993).			
RN	[5]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RX	MEDLINE=94090317; PubMed=826076;			
RA	Russell S.M., Kegan A.D., Harada N., Nakamura Y., Noguchi M.,			
RA	Leonard P., Friedmann M.C., Miyajima A., Puri R.K., Paul W.E.,			
RT	"Interleukin-2 receptor gamma chain: a functional component of the			
RT	interleukin-4 receptor.";			
RL	Science 262:1880-1883(1993).			
RN	[6]			
RP	IDENTIFICATION AS A IL-7R SUBUNIT.			

RX MEDLINE-94090316; PubMed-8266077;
 RA Noguchi M., Nakamura Y., Russell S.M., Ziegler S.F., Tsang M., Cao X.,
 RT Leonard W.J.;
 RT "Interleukin-2 receptor gamma chain: a functional component of the
 RT Interleukin-7 receptor.";
 RL Science 262:1877-1880(1993).
 RN [7]
 RP 3D-STRUCTURE MODELING OF 57-248.
 RX MEDLINE-95111955; PubMed-7529123;
 RA Bamorough P., Hedgecock C.J., Richards W.G.;
 RT "The interleukin-2 and interleukin-4 receptors studied by molecular
 RT modelling.";
 RL Structure 2:839-851(1994).
 RN [8]
 RP VARIANTS XSCID PHE-115, CYS-240 AND ILE-241.
 RX MEDLINE-94130970; PubMed-829698;
 RA Disanto J.P., Dautry-Varsat A., Certain S., Fischer A.,
 RA de Saint Basile G.;
 RT "Interleukin-2 (IL-2) receptor gamma chain mutations in X-linked
 RT severe combined immunodeficiency disease result in the loss of
 RT high affinity IL-2 receptor binding.";
 RL Eur. J. Immunol. 24:475-479(1994).
 RN [9]
 RP VARIANT XSCID IYS-68.
 RX MEDLINE-94375038; PubMed-8088810;
 RA Markiewicz S., Subtil A., Dautry-Varsat A., Fischer A.,
 RA de Saint Basile G.;
 RT "Detection of three nonsense mutations and one missense mutation in
 RT the Interleukin-2 receptor gamma chain gene in SCIDX1 that
 RT differently affect the mRNA processing.";
 RL Genomics 21:291-293(1994).
 RN [10]
 RP VARIANT XSCID HIS-162.
 RX MEDLINE-94300093; PubMed-8027558;
 RA Ishii N., Asao H., Kimura Y., Takeshita T., Nakamura M., Tsuchiya S.,
 RA Kono T., Meeda M., Uchihara Y., Sugamura K.;
 RT "Impairment of ligand binding and growth signaling of mutant IL-2
 RT receptor gamma-chains in patients with X-linked severe combined
 RT immunodeficiency.";
 RL J. Immunol. 153:1310-1317(1994).
 RN [11]
 RP VARIANT XSCID ASN-39.
 RX MEDLINE-95023932; PubMed-7937790;
 RA Disanto J.P., Rieux-Laucat F., Dautry-Varsat A., Fischer A.,
 RA de Saint Basile G.;
 RT "Defective human Interleukin 2 receptor gamma chain in an atypical X
 RT chromosome-linked severe combined immunodeficiency with peripheral T
 RT cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9466-9470(1994).
 RN [12]
 RP VARIANTS XSCID CYS-226 AND HIS-226.
 RX MEDLINE-95397841; PubMed-7668284;
 RA Pepper A.E., Buckley R.H., Small T.N., Puck J.M.;
 RT "Two mutational hotspots in the interleukin-2 receptor gamma chain
 RT gene causing human X-linked severe combined immunodeficiency.";
 RL Am. J. Hum. Genet. 57:564-571(1995).
 RN [13]
 RP VARIANT XSCID SER-183.
 RX MEDLINE-96013903; PubMed-7557965;
 RA Clark P.A., Lester T., Genet S., Jones A.M., Hendriks R.,
 RA Levinsky R.L., Kinon C.;
 RT "Screening for mutations causing X-linked severe combined
 RT immunodeficiency in the IL-2R gamma chain gene by single-strand
 RT conformation polymorphism analysis.";
 RL Hum. Genet. 96:427-432(1995).
 RN [14]
 RP VARIANT XSCID GLN-237 G-H-M INS.
 RX MEDLINE-95164726; PubMed-7860773;
 RA Puck J.M., Pepper A.E., Bedard P.-M., Laframboise R.;
 RT "Female germ line mosaicism as the origin of a unique IL-2 receptor
 RT gamma-chain mutation causing X-linked severe combined
 RT immunodeficiency.";
 RL J. Clin. Invest. 95:895-899(1995).

RN [15]
 RP VARIANT XSCID GLN-293.
 RX MEDLINE-95190013; PubMed-7883965;
 RA Schmaistleg F.C., Leonard W.J., Noguchi M., Berg M., Rudloff H.E.,
 RA Denney R.M., Dave S.K., Brooks E.G., Goldman A.S.;
 RT "Missense mutation in exon 7 of the common gamma chain gene causes a
 RT moderate form of X-linked combined immunodeficiency.";
 RL J. Clin. Invest. 95:1169-1173(1995).
 RN [16]
 RP VARIANT XSCID ARG-115.
 RX MEDLINE-97042245; PubMed-8900089;
 RA Stephan V., Mahn V., Le Delst F., Dirksen U., Broeker B.,
 RA Mueller-Fleckenstein I., Horneff G., Schroten H., Fischer A.,
 RA de Saint Basile G.;
 RT "Atypical X-linked severe combined immunodeficiency due to possible
 RT spontaneous reversion of the genetic defect in T cells.";
 RL New Engl. J. Med. 335:1563-1567(1996).
 RN [17]
 RP VARIANT XSCID GLN-285.
 RX MEDLINE-97295088; PubMed-9150740;
 RA Jones A.M., Clark P.A., Katz F., Genet S., McMahon C., Alterman L.,
 RA Cant A., Kinon C.;
 RT "B-cell-negative severe combined immunodeficiency associated with a
 RT common gamma chain mutation.";
 RL Hum. Genet. 99:677-680(1997).
 RN [18]
 RP VARIANT XSCID CYS-222.
 RX MEDLINE-98064061; PubMed-9399950;
 RA Sharfe N., Shahar M., Rolfman C.M.;
 RT "An interleukin-2 receptor gamma chain mutation with normal thymus
 RT morphology.";
 RL J. Clin. Invest. 100:3036-3043(1997).
 CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
 CC INTERLEUKINS.
 CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
 CC PROBABLY ALSO THE IL-13 RECEPTORS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A SEVERE COMBINED
 CC IMMUNODEFICIENCY, WHICH IS KNOWN AS AGAMMAGLOBULINEMIA, SWISS TYPE
 CC OR X-LINKED SEVERE COMBINED IMMUNODEFICIENCY DISEASE (XSCID).
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 CC -1- DATABASE: NAME=PROW: NOTE=X-CD guide CD132 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd132.htm".
 CC -1- DATABASE: NAME=IL2Gbase; NOTE=X-linked SCID mutation database;
 CC WWW="http://www.nhgrl.nih.gov/DIR/GMNB/SCID/".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D11086; BAA01857.1; -;
 DR EMBL; L12183; AAA59145.1; -;
 DR EMBL; L12178; AAA59145.1; JOINED.
 DR EMBL; L12176; AAA59145.1; JOINED.
 DR EMBL; L12177; AAA59145.1; JOINED.
 DR EMBL; L12179; AAA59145.1; JOINED.
 DR EMBL; L12180; AAA59145.1; JOINED.
 DR EMBL; L12181; AAA59145.1; JOINED.
 DR EMBL; L12182; AAA59145.1; JOINED.
 DR EMBL; L19546; AAC37524.1; -;
 DR PIR; A42565; A42565.
 DR PDB; 1ILM; 26-JAN-95.
 DR PDB; 1ILN; 26-JAN-95.
 DR MIM; 308380; -;
 DR MIM; 300400; -;
 DR InterPro; IPR002996; CR1A.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR003531; HemaTopo_receptor_S_F1.

Query Match 100.0%; Score 38; DB 1; Length 369;
 Best Local Similarity 100.0%; Pred. No. 0.57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
 |||||
 DB 120 KNIHLXQ 126

RESULT 2

COAT_CYWV STANDARD; PRT; 212 AA.
 AC P16486;
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DE Coat protein (Capsid protein).
 OS Clover yellow mosaic virus (CYMV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
 NCBI_TaxID=12177;
 OX NCBI_TaxID=12177;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89293092; PubMed=2738582;
 RA Abouhaidar M.G., Lai R.;
 RT "Nucleotide sequence of the 3'-terminal region of clover yellow
 mosaic virus RNA."
 CC J. Gen. Virol. 70:1871-1875(1989).
 CC -1- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS
 PARTICLES.
 CC -1- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: D00485; BAA00375.1; -
 DR InterPro: IPR000052; Potex_carlavirus_coat.
 DR Pfam: PF00286; virus_p-coat; 1.
 DR PRINTS: PR00232; POTXCARLCOAT.
 DR ProDom: PD00603; Potex_carlavirus_coat; 1.
 DR PROSITE: PS00418; POTEX_CARLAVIRUS_COAT; 1.
 KW Coat protein.
 SQ SEQUENCE 212 AA; 23438 MW; 30D6B8773F5D6E17 CRC64;

Query Match 84.2%; Score 32; DB 1; Length 212;
 Best Local Similarity 71.4%; Pred. No. 6.4;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
 |||||
 DB 171 KNIHLXQ 177

RESULT 3

YORO_TTV1 STANDARD; PRT; 51 AA.
 AC P19299;
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DE Hypothetical 6.2 kDa protein.
 OS Thermoplasma tenax virus 1 (strain KRA1) (TTV1).
 CC Viruses; dsDNA viruses, no RNA stage; Lipothirixviridae;
 CC Lipothirixvirus.
 OX NCBI_TaxID=10480;
 RN [1]
 RP SEQUENCE FROM N.A.
 SQ SEQUENCE FROM N.A.

RA Neumann H.;
 RL Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X14855; CAA32995.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 51 AA; 6171 MW; 7B2960FEFF195754 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 51;
 Best Local Similarity 83.3%; Pred. No. 2.3;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 EIHLYQ 7
 |||||
 DB 5 EIHLYQ 10

RESULT 4

PSB2_ORYSA STANDARD; PRT; 212 AA.
 AC Q9LST6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Proteasome subunit beta type 2 (EC 3.4.25.1) (20S proteasome alpha
 subunit D) (20S proteasome subunit beta-4).
 GN PBD1.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriaristaceae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RX MEDLINE=20314477; PubMed=10854779;
 RA Sassa H., Oguchi S., Inoue T., Hirano H.;
 RT "Primary structural features of the 20S proteasome subunits of rice
 (Oryza sativa)."
 RL Gene 250:61-66(2000).
 CC -1- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
 WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG,
 PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR
 SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC
 ACTIVITY (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad
 specificity.
 CC -1- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal
 proteolytic pathway.
 CC -1- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL
 SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE (BY
 SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B; ALSO KNOWN AS THE
 PROTEASOME B-TYPE FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AB026566; BAA96837.1; -

DR InterPro: IPR001353; Proteasome.
DR InterPro: IPR000243; Proteasome_B.
DR Pfam: PF00227; Proteasome; 1.
DR PROSITE: PS00854; PROTEASOME_B; FALSE_NEG.
KM Proteasome: Hydrolyase; Protease.
SO SEQUENCE 212 AA; 23444 MW; 9A87F8050078FF9 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 212;
Best Local Similarity 71.4%; Pred. No. 11;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KEHLQ 7
1 : 1111
Db 62 KNLHLQ 68

RESULT 5
ID PPNK_VIBCH STANDARD; PRT; 294 AA.
AC Q9KTP8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)
DE (Poly(P)/ATP NAD kinase).
GN PPNK OR VC0853.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Givann M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Esmolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberlack T., Fleischmann J.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae";
RT cholerase";
RL Nature 406:477-483(2000).
CC - FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes
CC ATP and other nucleoside triphosphates as well as inorganic
CC polyphosphate as a source of phosphorus (By similarity).
CC - CATALYTIC ACTIVITY: ATP + NAD(+) -> ADP + NADP(+).
CC - COFACTOR: Requires divalent metal ions for activity (By
CC similarity).
CC - SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC - SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL: AE004170; AAF94015.1; -
DR TIGR: VC0853; -
DR InterPro: IPR002504; DUF15.
DR Pfam: PF01513; DUF15; 1.
KM Transferase: Kinase; NAD: NADP: Complete proteome.
SO SEQUENCE 294 AA; 32698 MW; 7FEAA38920AAAE8 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 294;
Best Local Similarity 83.3%; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EHLQ 7
111:11
Db 259 EHLQ 264

RESULT 6
ID CYRG_CANFA STANDARD; PRT; 373 AA.
AC P40321;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cytokine receptor common gamma chain precursor (gamma-C)
DE (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).
GN IL2RG.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Plissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=95110114; PubMed=7829104;
RA Hentrich P.S., Somberg R.L., Fimiani V.M., Puck J.M., Patterson D.F.,
RA Felsburg P.J.;
RT "IL-2R gamma gene microdeletion demonstrates that canine X-linked
RT severe combined immunodeficiency is a homologue of the human
RT disease";
RL Genomics 23:69-74(1994).
CC - FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
CC INTERLEUKINS.
CC - SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC PROBABLY ALSO THE IL-13 RECEPTORS.
CC - SUBCELLULAR LOCATION: TYPE I membrane protein.
CC - DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A CANINE X-LINKED
CC SEVERE COMBINED IMMUNODEFICIENCY.
CC - SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC - SIMILARITY: CONTAINS 1 FIBROCTIN TYPE III-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL: U04361; AAC48403.1; -
DR HSSP: P31785; IILN.
DR InterPro: IPR002996; CRIA.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003531; Hematopo_receptor_s_fl.
DR Pfam: PF00041; fn3; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS01355; HEMATOPO_REC_S_FL; 1.
KM Receptor: Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 22
FT CHAIN 23 373
FT DOMAIN 23 261
FT TRANSSEM 262 283
FT DOMAIN 284 373
FT DOMAIN 151 249
FT DISULFID 62 72
FT DISULFID 102 115
FT CARBOHYD 24 24
FT CARBOHYD 71 71
FT CARBOHYD 75 75
FT CARBOHYD 84 84
FT CARBOHYD 159 159
FT CARBOHYD 164 164
FT CARBOHYD 249 249
FT SEQUENCE 373 AA; 42516 MW; 03A0DE1F8B089DBB CRC64;

Query Match 81.6%; Score 31; DB 1; Length 373;
 Best Local Similarity 71.4%; Pred. No. 19;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLVQ 7
 :|||||:
 Db 120 EETILYVE 126

RESULT 7

NU49_YEAST STANDARD; PRT; 472 AA.
 AC 002199;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Nucleoporin NUP49/NSP49 (Nuclear pore protein NUP49/NSP49).
 GN NUP49 OR NSP49 OR YGL172W OR G1648.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxId=4932;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA MEDLINE=9309880; PubMed=1464327;
 RA Wimmer C., Doye V., Grandi P., Nehrass U., Hurt E.C.:
 RT "A new subclass of nucleoporins that functionally interact with
 RT nuclear pore protein NSP1".
 RL EMOB J. 11:5051-5061(1992).
 RN [2]
 RN SEQUENCE FROM N.A.
 RA MEDLINE=93054906; PubMed=1385442;
 RA Wente S.R., Rout M.P., Blobel G.:
 RT "A new family of yeast nuclear pore complex proteins".
 RL J. Cell Biol. 119:703-723(1992).
 RN [3]
 RN SEQUENCE FROM N.A.
 RA STRAIN=5288C / FY1679;
 RA MEDLINE=96109931; PubMed=8619317;
 RA Bertani I., Coglievina M., Zaccaria P., Klita R., Bruschi C.V.:
 RT "The sequence of an 11.1 kb fragment on the left arm of Saccharomyces
 RT cerevisiae chromosome VII reveals six open reading frames including
 RT NSP49, KEM1 and four putative new genes".
 RL Yeast 11:1187-1194(1995).
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
 CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
 CC THE PROTEINS DURING NUCLEOTOPOLASMIC TRANSPORT.
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
 CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
 CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X68109; CAA48229.1; -
 DR EMBL: 215040; CAA78758.1; -
 DR EMBL: X84705; CAA59181.1; -
 DR EMBL: 272694; CAA96884.1; -
 DR PIR: S28026; S28026.
 DR PIR: S28538; S28538.
 DR PIR: C44402; C44402.
 DR SGD: S0003140; NUP49.
 KW Nuclear protein; Transport; Coiled coil; Repeat.
 FT DOMAIN 14 236 14 X 6 AA APPROXIMATE REPEATS OF
 FT G-L-F-G.
 FT COILED COIL (POTENTIAL).
 FT SEQUENCE 472 AA; 49142 MW; 0CAV516FF5753CA3 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 472;
 Best Local Similarity 71.4%; Pred. No. 25;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLVQ 7
 :|||||:
 Db 375 KTHLYVE 381

RESULT 8

CPSB_DROME STANDARD; PRT; 756 AA.
 AC 09V3D6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Probable cleavage and polyadenylation specificity factor, 100 kDa
 DE subunit (CPSF 100 kDa subunit).
 GN BCDNA:ID14168 OR CG1957.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Tsang G., Brokstein P., Frise E., Harvey D., Evans-Holm M.,
 RA Lewis S.E., Sun C., Rubin G.M.:
 RT "Full length Drosophila melanogaster cDNA sequence".
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBD databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RA STRAIN=BERKLEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortan J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazet R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abil J.F., Agbayani A., An H.-D., Andrews-Pfannkoch C., Baldwin D.,
 RA Baillif R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ijigawa C.,
 RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostrelli A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Syritskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
 RT "The genome sequence of Drosophila melanogaster".

RL Science 287:2185-2195(2000).
CC -1- FUNCTION: CPSF PLAYS A KEY ROLE IN PRE-MRNA 3'-END FORMATION,
CC RECOGNIZING THE AAUAAA SIGNAL SEQUENCE AND INTERACTING WITH
CC POLY(A) POLYMERASE AND OTHER FACTORS TO BRING ABOUT CLEAVAGE AND
CC POLY(A) ADDITION (BY similarity).
CC -1- SUBUNIT: CPSF IS A HETEROTETRAMER COMPOSED OF FOUR DISTINCT
CC SUBUNITS 160, 100, 70 AND 30 KDa (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE CPSF100 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF160933; AAD46873.1; -.
DR EMBL: AE003768; AAF56844.1; -.
DR FLYBASE: FBgn0027873; BcDNA: LD14168.
KW mRNA processing; Nuclear protein; RNA-binding.
SQ SEQUENCE 756 AA; 85418 MW; E391D61CFADDD821 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 756;
Best Local Similarity 83.3%; Pred. No. 41;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 EIHLYX 7
111111
Db 611 EIHLYX 616

RESULT 9
REPA_BPT4 STANDARD; PRT; 50 AA.
ID REPA_BPT4
AC P32284;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA replication protein repA (DNA-binding protein dbpB).
GN REPA OR DBP.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91190815; PubMed-2488704.
RA Mosig G., Lin G.W., Franklin J., Fan W.H.;
RT "Functional relationships and structural determinants of two
RT bacteriophage T4 lysozymes: a soluble (gene e) and a baseplate-
RT associated (gene 5) protein.";
RL New Biol. 1:171-179(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA Kuter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA Masyanzhinov V., Ruger W., Seidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION, AND FUNCTION
RX MEDLINE-20026826; PubMed-10559179;
RA Vaiskunnate R., Miller A., Davenport L., Mosig G.;
RT "Two new early bacteriophage T4 genes, repA and repB, that are
RT important for DNA replication initiated from origin E.";
RL J. Bacteriol. 181:715-7125(1999).
CC -1- FUNCTION: Involved in T4 DNA replication. Binds to ssDNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15728; -. NOT_ANNOTATED_CDS.
DR EMBL: AF158101; AAD42504.1; -.
KW DNA-binding; DNA replication.
SQ SEQUENCE 50 AA; 6130 MW; 33AA9EE42026EF6 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 50;
Best Local Similarity 83.3%; Pred. No. 3.7;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 KEIHLY 6
111111
Db 39 KDIHLY 44

RESULT 10
CLP1_PSEAE STANDARD; PRT; 213 AA.
ID CLP1_PSEAE
AC Q912U1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent Clp protease proteolytic subunit 1 (EC 3.4.21.92)
DE (Endopeptidase Clp 1).
GN CLP1 OR PA1801.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Medman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: Cleaves peptides in various proteins in a process that
CC requires ATP hydrolysis. Has a chymotrypsin-like activity. Plays a
CC major role in the degradation of misfolded proteins (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins to small peptides in
CC the presence of ATP and magnesium. Alpha-casein is the usual test
CC substrate. In the absence of ATP, only oligopeptides shorter than
CC five residues are cleaved (such as succinyl-leu-tyr-l-NMEC; and
CC leu-tyr-leu-l-tyr-tyr, in which the cleavage of the -Tyr-l-Leu-
CC and -Tyr-l-Tyr-bond also occurs).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S14; ALSO KNOWN AS CLPP
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE004606; AAC05190.1; -.
DR InterPro: IPR001907; Clp protease.
DR Pfam: PF00574; Clp protease; 1.
DR PRINTS: PR00127; CLPPROTEASEP.

DR PROSITE; PS00382; CLP_PROTEASE_HIS; 1.
 DR PROSITE; PS00381; CLP_PROTEASE_SER; 1.
 KM Hydrolase; Serine protease; Complete proteome.
 FT ACT_SITE 114 114 BY SIMILARITY.
 FT ACT_SITE 139 139 BY SIMILARITY.
 SQ SEQUENCE 213 AA; 23502 MW; ED5A73656E5A85E3 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 213;
 Best Local Similarity 83.3%; Pred. No. 18;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLX 6
 Db 74 KDHLX 79

RESULT 11

COAT_PMV STANDARD; PRT; 215 AA.
 ID COAT_PMV
 AC P16596;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Coat protein (Capsid protein).
 OS Papaya mosaic potexvirus (PMV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
 OX NCBI_TaxID=12181;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89381685; PubMed=2778435;
 RA Abouhalaid M.G.; Holy S.;
 RT "Nucleotide sequence of papaya mosaic virus RNA.";
 RL J. Gen. Virol. 70:2325-2331(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88089532; PubMed=3335832;
 RA Abouhalaid M.G.;
 RT "Nucleotide sequence of the capsid protein gene and 3' non-coding
 region of papaya mosaic virus RNA.";
 RL J. Gen. Virol. 69:219-226(1988).
 RN [3]
 RP SEQUENCE.
 RA Short M.N., Turner D.S., March J.F., Papin D.J.C., Parente A.,
 RA Davies J.W.;
 RT "The primary structure of papaya mosaic coat protein.";
 RL Virology 152:280-283(1986).
 CC -1- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS
 PARTICLES.
 CC -1- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-stb.ch).
 CC -----
 CC EMBL; D00240; BAA00169.1; -;
 DR EMBL; D13957; BAA03054.1; -;
 DR PIR; J00100; VCMGPM.
 DR InterPro: IPR000052; Potex.carlavirius_coat.
 DR Pfam: PF00286; virus_p-coat.1.
 DR PRINTS; PR00232; POTXCARLCOAT.
 DR ProDom: PD000603; Potex.carlavirius_coat; 1.
 DR PROSITE; PS00418; POTEX_CARLAVIRIUS_COAT; 1.
 KM Coat protein.
 FT .CONFLICT 1 5 MISSING (IN REF. 3).
 FT .CONFLICT 11 11 T -> I (IN REF. 1).
 FT .CONFLICT 30 30 S -> SN (IN REF. 3).
 FT .CONFLICT 77 77 MISSING (IN REF. 2).
 FT .CONFLICT 185 185 Q -> E (IN REF. 3).
 RN CONFLICT

FT CONFLICT 190 190 T -> A (IN REF. 3).
 FT CONFLICT 215 215 E -> Q (IN REF. 3).
 SQ SEQUENCE 215 AA; 23033 MW; 605B94EE70B41323 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 215;
 Best Local Similarity 57.1%; Pred. No. 18;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLX 7
 Db 175 KQVHLFQ 181

RESULT 12

HPPD_PSESP STANDARD; PRT; 357 AA.
 ID HPPD_PSESP
 AC P80064;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27) (4HPPD) (HPD).
 OS Pseudomonas sp. (Strain P.J. 874).
 OC Bacteria; Proteobacteria.
 OX NCBI_TaxID=306;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=92241278; PubMed=1572351;
 RA Ruetschi U., Odelhoeg B., Lindstedt S., Barros-Soederling J.,
 RA Persson B., Joernvall H.;
 RT "Characterization of 4-hydroxyphenylpyruvate dioxygenase. Primary
 structure of the Pseudomonas enzyme.";
 RL Eur. J. Biochem. 205:459-466(1992).
 CC -1- CATALYTIC ACTIVITY: 4-hydroxyphenylpyruvate + O(2) -> homogentisate
 + CO(2).
 CC -1- COFACTOR: IRON.
 CC -1- PATHWAY: CATABOLISM OF TYROSINE; SECOND STEP, CATABOLISM OF
 PHENYLALANINE; THIRD STEP.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SIMILARITY: BELONGS TO THE 4HPPD FAMILY.
 DR PIR; S21209; S21209.
 DR InterPro: IPR000325; Glyoxalase_1.
 DR Pfam: PF00903; Glyoxalase; 1.
 KM Oxidoreductase; Dioxygenase; Iron; Phenylalanine catabolism;
 KM Tyrosine catabolism.
 FT DOMAIN 167 196 TYR-RICH.
 SQ SEQUENCE 357 AA; 40060 MW; 26CF4A80B1484BD0 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 357;
 Best Local Similarity 57.1%; Pred. No. 31;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLX 7
 Db 44 KDVLX 50

RESULT 13
 DYHC_SCHPO STANDARD; PRT; 4196 AA.
 ID DYHC_SCHPO
 AC O13290; O9P6L0; O9UTP8;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dynein heavy chain, cytosolic (DYHC).
 GN DHCI OR SPAC30C2.01C OR SPAC1093.06C.
 OS Schizosaccharomyces pombe (Fission Yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=CRL152;
RA Yamamoto A., West R.R., McIntosh J.R., Hiraoka Y.;
RN Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE OF 1-2340 FROM N.A.
RC STRAIN=972;
RA Saunders D., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RN Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE OF 2308-4196 FROM N.A.
RC STRAIN=972;
RA Seeger K., Harris D., McDougall R.C., Rajandream M.A., Barrell B.G.;
RN Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RL [1]
CC FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES. REQUIRED FOR NUCLEAR MOVEMENT
CC DURING MEIOTIC PROPHASE.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. PROBABLY BINDS INDIRECTLY TO
CC THE INNER PLASMA MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB006784; BAA22056.1; -
DR EMBL: AL353652; CAB90788.1; -
DR EMBL: AL132839; CAB60251.1; -
KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT DOMAIN 1217 1252 COILED COIL (POTENTIAL).
FT DOMAIN 1984 2012 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 3315 3403 COILED COIL (POTENTIAL).
FT DOMAIN 3649 3666 COILED COIL (POTENTIAL).
FT NP_BIND 1890 1897 ATP (POTENTIAL).
FT NP_BIND 2169 2176 ATP (POTENTIAL).
FT NP_BIND 2174 2181 ATP (POTENTIAL).
FT NP_BIND 2520 2527 ATP (POTENTIAL).
SQ SEQUENCE 4196 AA; 484308 MW; 8F10AE370184FC0C CRC64;

Query Match 78.9%; Score 30; DB 1; Length 4196;
Best Local Similarity 71.4%; Pred. No. 4,3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLXQ 7
Db 1320 EEIHLXK 1326

RESULT 14
YGBS_PASMU STANDARD; PRT; 233 AA.
AC O9CKD7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein PM1685.
GN PM1685.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;

```

```

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- SIMILARITY: BELONGS TO THE PIRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE006205; AAK03769.1; -
DR InterPro: IPR003829; DUF209.
DR Pfam: PF02678; DUF209; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 233 AA; 26804 MW; 28500A6C859F65D5 CRC64;

Query Match 76.3%; Score 29; DB 1; Length 233;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 EIHLYQ 7
Db 113 DVHLYQ 118

RESULT 15
YGBS_EDWIC STANDARD; PRT; 286 AA.
AC O52401;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 30.6 kDa protein in IC1A-FBA intergenic region.
GN YGBS.
OS Edwardsiella ictaluri.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Edwardsiella.
OX NCBI_TaxID=67780;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=93-146;
RA Moore M.M., Fernandez D.H., Thune R.L.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0003 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF037440; AAB92571.1; -
DR InterPro: IPR001880; UPF0003.
DR Pfam: PF00924; MS channel; 1.
DR PROSITE: PS01246; UPF0003; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
SQ SEQUENCE 286 AA; 30648 MW; DB2442E07C70835D CRC64;

Query Match 76.3%; Score 29; DB 1; Length 286;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

OY 2 EIHLYQ 7
: : : : :
Db 274 DVHLYQ 279

Search completed: August 20, 2002, 11:33:01
Job time: 1450 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:31:53 ; Search time 191.14 Seconds
(without alignments)
6.335 Million cell updates/sec

Title: US-09-824-286-14
Perfect score: 38
Sequence: 1 KEIHLXQ 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	86.8	1611	10 Q9LKW5	Q9LKW5 aegilops ta
2	33	86.8	1628	10 Q9LKW6	Q9LKW6 triticum ae
3	32	84.2	257	12 Q6K343	Q6K343 clover yell
4	32	84.2	275	4 Q9H285	Q9H285 homo sapien
5	32	84.2	286	12 Q6K342	Q6K342 clover yell
6	32	84.2	312	12 Q6K339	Q6K339 clover yell
7	32	84.2	367	4 Q9GELA	Q9GELA homo sapien
8	32	84.2	434	2 Q9JIS0	Q9JIS0 salmonella
9	32	84.2	551	10 Q9SVB6	Q9SVB6 arabidopsi
10	31	81.6	184	5 Q7J159	Q7J159 tetrahymena
11	31	81.6	233	16 P7J155	P7J155 synecocyst
12	31	81.6	506	10 Q9FHY9	Q9FHY9 arabidopsi
13	31	81.6	591	5 Q17412	Q17412 caenorhabdi
14	31	81.6	634	16 Q9GNH8	Q9GNH8 pasteurella
15	31	81.6	673	16 Q07549	Q07549 bacillus su
16	31	81.6	677	5 Q9VEL0	Q9VEL0 drosophila

17	31	81.6	785	5 Q9V089	Q9V089 drosophila
18	31	81.6	867	5 Q95RC2	Q95RC2 drosophila
19	30	78.9	135	16 Q97PW3	Q97PW3 streptococ
20	30	78.9	188	16 P72984	P72984 synecocyst
21	30	78.9	191	2 Q9AKB0	Q9AKB0 rickettsia
22	30	78.9	223	16 Q9CHA8	Q9CHA8 lactococcus
23	30	78.9	244	2 Q34250	Q34250 wolfinella s
24	30	78.9	255	12 Q919K4	Q919K4 culix nigri
25	30	78.9	292	16 Q989S3	Q989S3 rhizobium l
26	30	78.9	319	17 Q97UJ8	Q97UJ8 sulfolobus
27	30	78.9	433	2 P94847	P94847 helicobacte
28	30	78.9	433	16 Q25321	Q25321 helicobacte
29	30	78.9	433	16 Q9ZLN2	Q9ZLN2 helicobacte
30	30	78.9	456	3 Q9US58	Q9US58 schizosach
31	30	78.9	480	10 Q9M5C4	Q9M5C4 euphorbia e
32	30	78.9	482	5 Q9U0J4	Q9U0J4 plasmodium
33	30	78.9	553	5 Q9NFK4	Q9NFK4 boophilus m
34	30	78.9	553	5 Q9NFK3	Q9NFK3 boophilus m
35	30	78.9	553	5 Q9NFK2	Q9NFK2 boophilus m
36	30	78.9	553	5 Q9NFK1	Q9NFK1 boophilus m
37	30	78.9	752	12 Q913V5	Q913V5 baboon gamm
38	30	78.9	752	2 P96793	P96793 lactobacill
39	30	78.9	833	3 Q96V83	Q96V83 metarhizium
40	30	78.9	940	16 Q9KR83	Q9KR83 vibrio chol
41	30	78.9	1220	15 Q41894	Q41894 bovine sync
42	30	78.9	1322	10 Q9PLN3	Q9PLN3 arabidopsi
43	29	76.3	129	16 Q9PEU1	Q9PEU1 xyella fas
44	29	76.3	146	5 Q18058	Q18058 caenorhabdi
45	29	76.3	153	16 Q25338	Q25338 helicobacte

ALIGNMENTS

RESULT 1	
Q9LKW5	PRELIMINARY; PRT; 1611 AA.
AC Q9LKW5	
DT 01-OCT-2000 (TREMBLrel. 15, Created)	
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)	
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE STARCH SYNTHASE III.	
OS Aegilops tauschii (Aegilops squarrosa).	
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;	
OC Triticeae; Aegilops.	
OX NCBI_TaxID=37682;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=20317203; PubMed=10859191;	
RA Li Z., Mouille G., Kosar-Hashemi B., Rahman S., Clarke B., Gale K.R.,	
RA Appels R., Morell M.K.;	
RT "The structure and expression of the wheat starch synthase III gene.	
RT Motifs in the expressed gene define the lineage of the starch synthase	
RT III gene family.";	
RL Plant Physiol. 123:613-624(2000).	
DR EMBL: AF258609; AAF88000.1; "	
SQ SEQUENCE 1611 AA; 180653 MW; 78BEC99C03FEAF9 CRC64;	

Query Match 86.8%; Score 33; DB 10; Length 1611;
Best Local Similarity 71.4%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY	1 KEIHLXQ 7
DB	1215 KDLHLXQ 1221
RESULT 2	
Q9LKW6	PRELIMINARY; PRT; 1628 AA.
ID Q9LKW6	
AC Q9LKW6;	

DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE STARCH SYNTHASE III.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
FX MEDLINE=20317203; PubMed=10859191;
RA Li Z., Mouille G., Kosar-Hashemi B., Rahman S., Clarke B., Gale K.R.,
RA Appels R., Morell M.K.;
RT "The structure and expression of the wheat starch synthase III gene.
RT Motifs in the expressed gene define the lineage of the starch synthase
RT II gene family.";
RL Plant Physiol. 123:613-624(2000).
DR EMBL: AF258608; AAF87999.1;
SQ SEQUENCE 1628 AA; 183138 MW; 069498ADA859B819 CRC64;

Query Match 86.8%; Score 33; DB 10; Length 1628;
Best Local Similarity 71.4%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEIHLQ 7
Db 1232 KDLHLQ 1238

RESULT 3
066343 PRELIMINARY; PRT; 257 AA.
AC 066343;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE COAT PROTEIN (CAPSID PROTEIN).
OS Clover yellow mosaic virus (CYMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
NCBI_TaxID=12177;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89293092; PubMed=2738582;
RA Aboumehar M.G., Lai R.;
RT "Nucleotide sequence of the 3'-terminal region of clover yellow mosaic
RT virus RNA.";
RL J. Gen. Virol. 70:1871-1875(1989).
CC -1- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS PARTICLES
CC (BY SIMILARITY).
CC -1- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.
DR EMBL: D00485; BAA00373.1; -
DR InterPro: IPR000052; Potex_carlaviruses_coat.
DR InterPro: IPR000531; TONB_boxC.
DR Pfam: PF00286; virus_P-coat; 1.
DR PRINTS: PR00232; POTXCARLCOAT.
DR ProDom: PD000603; Potex_carlavirus_coat; 1.
DR PROSITE: PS00418; POTEX_CARLAVIRUS_COAT; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KM Coat protein.
SQ SEQUENCE 257 AA; 28095 MW; 4D9FA346DE9CACC6 CRC64;

Query Match 84.2%; Score 32; DB 12; Length 257;
Best Local Similarity 71.4%; Pred. No. 47;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KEIHLQ 7
Db 216 KNVHLQ 222

RESULT 4
09H285 PRELIMINARY; PRT; 275 AA.
AC 09H285;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SEROLOGICALLY DEFINED BREAST CANCER ANTIGEN NY-BR-24 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BREAST;
RA Scanlan M.J., Gout I., Stockert E., Gure A.O., Jaeger D., Chen Y.-T.,
RA Old L.J.;
RT "Humoral Immunity to Human Breast Cancer: Antigen Definition and
RT Quantitative Analysis of mRNA Expression.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF308288; AAC48256.1; -
DR InterPro: IPR000038; GTP_Cell_Div.
DR Pfam: PF00735; GTP_CDC; 1.
DR ProDom: PD002565; GTP_Cell_Div; 1.
FT NON_TER 1
SQ SEQUENCE 275 AA; 31893 MW; 517068962B98C7B1 CRC64;

Query Match 84.2%; Score 32; DB 4; Length 275;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEIHLQ 7
Db 102 EIHLYQ 108

RESULT 5
066342 PRELIMINARY; PRT; 286 AA.
AC 066342;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE COAT PROTEIN (CAPSID PROTEIN).
OS Clover yellow mosaic virus (CYMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
NCBI_TaxID=12177;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91306431; PubMed=1830181;
RA White K.A., Bancroft J.B., Mackie G.A.;
RT "Defective RNAs of clover yellow mosaic virus encode
RT nonstructural/coat protein fusion products.";
RL Virology 183:479-486(1991).
CC -1- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS PARTICLES
CC (BY SIMILARITY).
CC -1- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.
DR EMBL: M63514; AAA42938.1; -
DR InterPro: IPR000052; Potex_carlaviruses_coat.
DR Pfam: PF00286; virus_P-coat; 1.
DR ProDom: PD000603; Potex_carlavirus_coat; 1.
KM Coat protein.
SQ SEQUENCE 286 AA; 31730 MW; 9BE5B01BD1E32895 CRC64;

Query Match 84.2%; Score 32; DB 12; Length 286;
Best Local Similarity 71.4%; Pred. No. 52;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KEIHLQ 7
Db 245 KNVHLQ 251


```

RESULT 6
ID 066339 PRELIMINARY; PRT; 312 AA.
AC 066339;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COAT PROTEIN (CAPSID PROTEIN).
OS Clover yellow mosaic virus (CYMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Potexvirus.
NCBI_TaxID=12171;
RX MEDLINE=91306431; PubMed=1830181;
RA White K.A., Bancroft J.B., Mackie G.A.;
RT "Defective RNAs of clover yellow mosaic virus encode
nonstructural/coat protein fusion products."
RL Virology 183:479-486(1991).
CC -1- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS PARTICLES
(BY SIMILARITY).
CC -1- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.
DR EMBL: M63511; AAA2935.1; -
DR Interpro: IPR000052; Potex_carlavirius_coat.
DR Pfam: PF00286; Virus_P-coat; 1.
DR ProDom: PD000603; Potex_carlavirius_coat; 1.
DR ProSITE: PS00418; POTEX_CARLAVIRIUS_COAT; 1.
KW Coat protein.
SQ SEQUENCE 312 AA; 34657 MW; BA17BDFB3E675E21 CRC64;

```

```

Query Match
Best Local Similarity 84.2%; Score 32; DB 12; Length 312;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KEIHLVY 7
I :||||
DB 271 KNVHLVY 277

```

```

RESULT 7
ID 096EL4 PRELIMINARY; PRT; 367 AA.
AC 096EL4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SIMILAR TO SEPTIN 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PRIMARY B-CELLS FROM TONSILS;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC012161; AAH12161.1; -
SQ SEQUENCE 367 AA; 41970 MW; E9CGBEF373CB0FC4 CRC64;

```

```

Query Match
Best Local Similarity 84.2%; Score 32; DB 4; Length 367;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEIHLVY 7
I :|||||
DB 194 ERIHIYQ 200

```

```

RESULT 8
Q93ISO

```

```

ID Q93ISO PRELIMINARY; PRT; 434 AA.
AC Q93ISO;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SCIP PROTEIN.
OS Salmonella enterica subsp. enterica serovar Typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
NCBI_TaxID=90371;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SR-11;
RA Folkesson A.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SR-11;
RX MEDLINE=99348391; PubMed=10417651;
RA Folkesson A., Advani A., Sukupolvi S., Pfeiffer J.D., Normark S.,
Loefbehl S.;
RT "Multiple insertions of fimbrial operons correlate with the evolution
of Salmonella serovars responsible for human disease."
RL M01. Microbiol. 33:612-622(1999).
DR EMBL: AJ320483; CAC48213.1; -
SQ SEQUENCE 434 AA; 47355 MW; 3D787A09B8A092D CRC64;

```

```

Query Match
Best Local Similarity 84.2%; Score 32; DB 2; Length 434;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEIHLVY 7
I :|||||
DB 96 REMHLVY 102

```

```

RESULT 9
ID Q9SVE6 PRELIMINARY; PRT; 551 AA.
AC Q9SVE6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE GROWTH REGULATOR PROTEIN.
GN F22113.160 OR A14G38390.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Medler H., Kutzner M., Wambutt R., Bancroft I., Mewes H.W.,
Mayer K.F.X., Schueller C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Wedler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K.,
Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL035539; CAB37495.1; -
DR EMBL: AL161593; CAB80504.1; -
DR Interpro: IPR004348; DUF246.

```

DR Pfam; PF03138; DUF246; 1.
SQ SEQUENCE 551 AA; 62872 MW; 61FC1B84BA3756FC CRC64;

Query Match
Best Local Similarity 84.2%; Score 32; DB 10; Length 551;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLXQ 6
111111
Db 360 KEVHLX 365

RESULT 10
077159 PRELIMINARY; PRT; 184 AA.

AC 077159;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HETEROCHROMATIN-ASSOCIATED PROTEIN 1-LIKE PROTEIN.
GN HHP1.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_Taxid=5911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CU428;
RX MEDLINE-99030619; Pubmed-9811850;
RA Huang H., Willey E.A., Lending C.R., Allis C.D.;
RT "An HPI-like protein is missing from transcriptionally silent
RT micronuclei of Tetrahymena."
RL Proc. Natl. Acad. Sci. U.S.A. 95:13624-13629(1998).
DR EMBL; AF079405; AAC78328.1;
DR InterPro; IPR000953; Chromo.
DR Pfam; PF00385; Chromo. 1.
DR SMART; SM00298; CHROMO_2; 1.
DR PROSITE; PSS0013; CHROMO_2; 1.
SQ SEQUENCE 184 AA; 21140 MW; 8B3BD646B09CD39F CRC64;

Query Match
Best Local Similarity 81.6%; Score 31; DB 5; Length 184;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KEIHLXQ 7
111111
Db 115 KEIHNYQ 121

RESULT 11
073155 PRELIMINARY; PRT; 233 AA.

AC 073155;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)
DE HYPOHETICAL 26.1 KDA PROTEIN.
GN SLR1039.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_Taxid=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97061201; Pubmed-8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Suglura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR EMBL; D90904; BAA17181.1;
DR InterPro; IPR001601; Meth-transf.
DR InterPro; IPR000051; SAM_bind
DR InterPro; IPR004033; ubiE_COQ5_methyltransf.
DR Pfam; PF01209; ubiE_methyltran; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 233 AA; 26104 MW; 681F5B03AABD288 CRC64;

Query Match
Best Local Similarity 81.6%; Score 31; DB 16; Length 233;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 EIHLXQ 7
111111
Db 208 ELHLXQ 213

RESULT 12
09FHY9 PRELIMINARY; PRT; 506 AA.

AC 09FHY9;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE GTPASE ACTIVATOR PROTEIN OF RAB-LIKE SMALL GTPASES-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE-99397451; Pubmed-10470850;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT pl and TAC clones."
RL DNA Res. 6:183-195(1999).
DR EMBL; AB017067; BAB08427.1;
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00164; TBC; 1.
SQ SEQUENCE 506 AA; 57954 MW; 3132E97A118C5D5 CRC64;

Query Match
Best Local Similarity 81.6%; Score 31; DB 10; Length 506;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLXQ 7
111111
Db 377 KDHLX 383

RESULT 13
017412 PRELIMINARY; PRT; 591 AA.

AC 017412;
DT 01-JAN-1999 (TREMblrel. 09, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE B0001.3 PROTEIN.
GN B0001.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]

RA SEQUENCE FROM N.A.
RA Sims M.A.;
RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode *C.elegans*: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL: 269634; CAA93452.2;
SQ SEQUENCE 591 AA; 67490 MW; 0B4BA802581D0393 CRC64;

Query Match
Best Local Similarity 57.1%; Score 31; DB 5; Length 591;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
DB 214 KDVLXQ 220
1:1111:
1:1111:

RESULT 14
O9CNH8 PRELIMINARY; PRT; 634 AA.
AC O9CNH8;
ID 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHEICAL PROTEIN PM0453.
GN PM0453.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RA MEDLINE=21145866; PubMed=11248100;
RX May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genome sequence of *Pasteurella multocida* pm70."
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL: AE006081; AK02537.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 634 AA; 72154 MW; BE521565BC5145C8 CRC64;

Query Match
Best Local Similarity 71.4%; Score 31; DB 16; Length 634;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
DB 41 KTVHLXQ 47
1:1111:
41:KTVHLXQ 47

RESULT 15
O07549 PRELIMINARY; PRT; 673 AA.
AC O07549;
ID 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHEICAL 76.3 KDA PROTEIN.
GN YHEH.
OS *Bacillus subtilis*.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;

RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessières P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schoefer R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viara A., Wambutt R., Wedler E., Wedler H., Weltenegeger T.,
RA Winiers P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium *Bacillus*
RT *subtilis*."
RT Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y14080; CAA74449.1;
DR EMBL: Z99109; CAB12811.1;
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC_transporter_tmem.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane_1.
DR Pfam: PF00005; ABC_tran; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 673 AA; 76304 MW; 533DC9970E9B8D5 CRC64;

Query Match
Best Local Similarity 71.4%; Score 31; DB 16; Length 673;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
DB 86 KEAHLXQ 92
1:1111:
86:KEAHLXQ 92

Search completed: August 20, 2002, 11:31:55
Job time: 1469 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:07:16 ; Search time 187.18 Seconds
(without alignments)
4.154 Million cell updates/sec

Title: US-09-824-286-15

Perfect score: 34

Sequence: 1 LQNLVIP 7

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_032802.*
1: /SID5/gcgdata/geneseq/genesqp-emb1/AA1980.DAT.*
2: /SID5/gcgdata/geneseq/genesqp-emb1/AA1981.DAT.*
3: /SID5/gcgdata/geneseq/genesqp-emb1/AA1982.DAT.*
4: /SID5/gcgdata/geneseq/genesqp-emb1/AA1983.DAT.*
5: /SID5/gcgdata/geneseq/genesqp-emb1/AA1984.DAT.*
6: /SID5/gcgdata/geneseq/genesqp-emb1/AA1985.DAT.*
7: /SID5/gcgdata/geneseq/genesqp-emb1/AA1986.DAT.*
8: /SID5/gcgdata/geneseq/genesqp-emb1/AA1987.DAT.*
9: /SID5/gcgdata/geneseq/genesqp-emb1/AA1988.DAT.*
10: /SID5/gcgdata/geneseq/genesqp-emb1/AA1989.DAT.*
11: /SID5/gcgdata/geneseq/genesqp-emb1/AA1990.DAT.*
12: /SID5/gcgdata/geneseq/genesqp-emb1/AA1991.DAT.*
13: /SID5/gcgdata/geneseq/genesqp-emb1/AA1992.DAT.*
14: /SID5/gcgdata/geneseq/genesqp-emb1/AA1993.DAT.*
15: /SID5/gcgdata/geneseq/genesqp-emb1/AA1994.DAT.*
16: /SID5/gcgdata/geneseq/genesqp-emb1/AA1995.DAT.*
17: /SID5/gcgdata/geneseq/genesqp-emb1/AA1996.DAT.*
18: /SID5/gcgdata/geneseq/genesqp-emb1/AA1997.DAT.*
19: /SID5/gcgdata/geneseq/genesqp-emb1/AA1998.DAT.*
20: /SID5/gcgdata/geneseq/genesqp-emb1/AA1999.DAT.*
21: /SID5/gcgdata/geneseq/genesqp-emb1/AA2000.DAT.*
22: /SID5/gcgdata/geneseq/genesqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	7	AAW31652	Human cytokine rec
2	34	100.0	230	AAAR47151	IL-2 receptor gamma
3	34	100.0	230	AAAR82934	Interleukin 4 comp
4	34	100.0	252	AAAR47150	IL-2 receptor gamma
5	34	100.0	347	AAAR47149	IL-2 receptor gamma
6	34	100.0	369	AAAR47148	IL-2 receptor gamma
7	34	100.0	369	AAAR59094	Murine IL-2R gamma
8	34	100.0	482	AAW31646	Human cytokine rec
9	34	100.0	691	AAV92202	Fusion polypeptide
10	34	100.0	694	AAV92201	Fusion polypeptide
11	34	100.0	694	AAV92203	Fusion polypeptide

12	31	91.2	1374	19	AAW72225	HSV-2 strain SB5 C
13	31	91.2	1374	19	AAW69753	Herpes simplex vir
14	31	91.2	1384	19	AAW72224	HSV-2 strain SB5 C
15	31	91.2	1396	19	AAW72117	HSV-2 strain SB5 C
16	31	91.2	1396	19	AAW72039	HSV-2 strain SB5 C
17	30	88.2	81	22	ABBI6652	Human nervous syst
18	30	88.2	303	21	AAV77129	Marburg virus stru
19	30	88.2	553	21	AAV97004	S. cerevisiae esse
20	30	88.2	634	22	ABR60004	Drosophila melanog
21	29	85.3	125	22	AAV90109	C glutamicum prote
22	29	85.3	232	21	AAV50179	Arabidopsis thalia
23	29	85.3	246	22	AAU37393	Staphylococcus aur
24	29	85.3	246	22	AAU37447	Staphylococcus aur
25	29	85.3	315	21	AAV50193	Arabidopsis thalia
26	29	85.3	363	22	AAV71682	Bos taurus Interle
27	29	85.3	454	21	AAV50178	Arabidopsis thalia
28	29	85.3	537	21	AAV50192	Arabidopsis thalia
29	29	85.3	623	19	AAW75773	Arabidopsis thalia
30	29	85.3	625	22	AAU02044	Amino acid sequenc
31	29	85.3	633	22	AAV39511	B. thuringiensis t
32	29	85.3	643	22	AAW41297	Human polypeptide
33	28	82.4	54	22	AAW06181	Human polypeptide
34	28	82.4	73	21	AAV58387	Peptide #4863 enco
35	28	82.4	73	22	AAV74277	Lung cancer associ
36	28	82.4	83	22	AAV75671	Human colon cancer
37	28	82.4	84	17	AAV99710	Human colon cancer
38	28	82.4	86	21	AAV4184	Human cancer assoc
39	28	82.4	103	22	AAV62532	B. melitensis vitu
40	28	82.4	192	13	AAV2953	Promoter-secretion
41	28	82.4	192	14	AAV7660	Promoter-secretion
42	28	82.4	244	20	AAV37743	Amino acid sequenc
43	28	82.4	257	21	AAV36350	Arabidopsis thalia
44	28	82.4	263	21	AAV36349	Arabidopsis thalia
45	28	82.4	269	21	AAV36348	Arabidopsis thalia

ALIGNMENTS

RESULT 1

ID AAW31652 standard; Peptide; 7 AA.

AC AAW31652;

DF 21-MAY-1998 (first entry)

DE Human cytokine receptor gc chain epitope.

Cytokine receptor; gamma common chain; gc chain; human;
blocking agent; monoclonal antibody; CP-B8; immunological disease;
myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
insulin-dependent diabetes; inflammatory bowel disease;
sympathetic ophthalmia; uveitis; allergy; asthma; infection;
graft versus host disease; psoriasis; immunosuppressive; therapy;
epitope.

OS Homo sapiens.

PN WO9743416-A1.

PD 20-NOV-1997.

PF 09-MAY-1997; 97WO-US07870.

PR 10-MAY-1996; 96US-0017466.

PA (BIO) BIOGEN INC.

PI Benjamin CD, Burky LC, Hession C, Whitty A;

DR WPI; 1998-008885/01.

PT Blocking agents of the gamma common chain of cytokine receptors -
PT particularly monoclonal antibodies, used to induce T cell anergy for
PT treatment of immunological diseases
XX
PS Claim 24; Page 84; 111pp; English.
XX
CC This peptide comprises an epitope of the human cytokine receptor
CC common gamma (gc) chain (see AAW31646) that is recognised by
CC gc blocking agents of the invention. 5 Such epitopes (see
CC AAW31650-54) have been identified. The invention provides
CC compositions and methods for inhibiting cytokine signalling using
CC gc chain blocking agents for the treatment of immunological
CC diseases such as myasthenia gravis, rheumatoid arthritis, lupus,
CC multiple sclerosis, insulin-dependent diabetes, inflammatory bowel
CC disease, sympathetic ophthalmia, uveitis, allergy, asthma,
CC parasitic infection, graft vs. host disease or psoriasis. A
CC preferred gc blocking agent is Mab CP.B8 or its Fab fragment (see
CC also AAW31647-48).
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 34; DB 19; Length 7;
Best Local Similarity 100.0%; Pred. No. 6,4e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
D5 1 lqnlvip 7
|||||
1 lqnlvip 7

RESULT 2
AAW47151
ID AAR47151 standard; Protein: 230 AA.
XX
XX AAR47151:
AC
XX
XX 13-JUN-1994 (first entry)
DT
XX
XX IL-2 receptor gamma chain.
DE
XX
XX Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KM rheumatoid arthritis; transplant rejection; primer;
KW polymerase chain reaction; PCR; amplification.
XX
XX Homo sapiens.
OS
XX
XX EP578932-A.
PN
XX
XX 19-JAN-1994.
PD
XX
XX 22-APR-1993; 93EP-0106561.
PE
XX
XX 23-APR-1992; 92JP-0104947.
PR
XX
XX (AJIN) AJINOMOTO KK.
PA (SUGA/) SUGAMURA K.
XX
XX Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
XX WPI: 1994-017546/03.
DR N-PSDB; AA054831.
XX
XX DNA and protein sequences of IL-2 gamma chain - useful as immune
PT regulatory agents for treatment of e.g. rheumatoid arthritis and
PT transplant rejection
XX
XX Disclosure; Page 22-23, 35-36; 50pp; English.
PS
XX
XX The human IL-2 receptor gamma chain preform (AAR47148), including the
CC signal peptide, is encoded by the sequence given in AA054828. The
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
CC

CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
CC while a soluble form suitable for expression in prokaryotes (AAR47151)
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
CC terminal sequence of IL-2 receptor gamma chain, and are used to
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
CC are used to obtain the protein given in AAR47151.
XX
SQ Sequence 230 AA;

Query Match 100.0%; Score 34; DB 15; Length 230;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
D5 126 lqnlvip 132
|||||
126 lqnlvip 132

RESULT 3
AAR82934
ID AAR82934 standard; Protein: 230 AA.
XX
XX AAR82934:
AC
XX
XX 26-FEB-1996 (first entry)
DT
XX
XX Interleukin 4 component common to the IL-2 receptor gamma chain.
DE
XX
XX Interleukin-4; IL-4; gamma chain component; immunosuppressants;
KW anti-allergy agent; signal transduction inhibitor; atcoimmune;
KW disease; anti-inflammatory; anaphylactic shock; bronchial asthma;
KW Interleukin-2; IL-2; atopic dermatitis; urticaria.
XX
XX Homo sapiens.
OS
XX
XX JP07149662-A.
PN
XX
XX 13-JUN-1995.
PD
XX
XX 07-SEP-1994; 94JP-0213706.
PE
XX
XX 08-SEP-1993; 93JP-0223574.
PR
XX
XX (AJIN) AJINOMOTO KK.
PA (SUGA/) SUGAMURA K.
XX
XX WPI: 1995-243601/32.
DR N-PSDB; AAT04952.
XX
XX Novel interleukin-4 receptor monoclonal antibodies inhibit signal
PT transmission - useful as immunosuppressants and anti-allergy agents.
PT
XX
XX Example 1; Page 9; 11pp; Japanese.
PS
XX
XX AAT04952 encodes AAR82934 a component of the IL-4 receptor common to
CC the IL-2 receptor gamma chain molecule, which was used to generate
CC anti-IL-4 receptor monoclonal antibodies (mAbs). The mAbs (IL-4
CC signal transduction inhibitors) can be used as immunosuppressants
CC and anti-allergy agents, for the treatment of autoimmune and chronic
CC inflammatory diseases, e.g. anaphylactic shock, bronchial asthma,
CC atopic dermatitis and urticaria.
XX
SQ Sequence 230 AA;

Query Match 100.0%; Score 34; DB 16; Length 230;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
D5 127 lqnlvip 133
|||||
127 lqnlvip 133

```

RESULT 4
ID AAR47150 standard; Protein; 252 AA.
XX
AC AAR47150;
XX
DT 13-JUN-1994 (first entry)
XX
DE IL-2 receptor gamma chain.
XX
KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KW rheumatoid arthritis; transplant rejection; primer;
KW polymerase chain reaction; PCR; amplification; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= Sig_peptide
XX
PN EP578932-A.
XX
PD 19-JAN-1994.
XX
PE 22-APR-1993; 93EP-0106561.
XX
PR 23-APR-1992; 92JP-0104947.
XX
XX (AJIN ) AJINOMOTO KK.
XX (SUGA/) SUGAMURA K.
XX
PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
DR WPI; 1994-017546/03.
DR N-PSDB; AA054830.
XX
PT DNA and protein sequences of IL-2 gamma chain - useful as immune
PT regulatory agents for treatment of e.g. rheumatoid arthritis and
PT transplant rejection
XX
PS Disclosure; Page 21-22, 34-35; 50pp; English.
XX
CC The human IL-2 receptor gamma chain preform (AAR47148), including the
CC signal peptide, is encoded by the sequence given in AA054828. The
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
CC while a soluble form suitable for expression in prokaryotes (AAR47151)
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
CC terminal sequence of IL-2 receptor gamma chain, and are used to
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
CC are used to obtain the protein given in AAR47151.
XX
SQ Sequence 252 AA;

```

```

Query Match 100.0%; Score 34; DB 15; Length 252;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 LQNLVIP 7
DB 148. lqnlvip 154

```

```

RESULT 5
ID AAR47149 standard; Protein; 347 AA.
XX
AC AAR47149;
XX

```

```

DT 13-JUN-1994 (first entry)
XX
DE IL-2 receptor gamma chain.
XX
KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KW rheumatoid arthritis; transplant rejection; primer;
KW polymerase chain reaction; PCR; amplification.
XX
OS Homo sapiens.
XX
PN EP578932-A.
XX
PD 19-JAN-1994.
XX
PE 22-APR-1993; 93EP-0106561.
XX
PR 23-APR-1992; 92JP-0104947.
XX
XX (AJIN ) AJINOMOTO KK.
XX (SUGA/) SUGAMURA K.
XX
PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
DR WPI; 1994-017546/03.
DR N-PSDB; AA054829.
XX
PT DNA and protein sequences of IL-2 gamma chain - useful as immune
PT regulatory agents for treatment of e.g. rheumatoid arthritis and
PT transplant rejection
XX
PS Claim 4; Page 41; 50pp; English.
XX
CC The human IL-2 receptor gamma chain preform (AAR47148), including the
CC signal peptide, is encoded by the sequence given in AA054828. The
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
CC while a soluble form suitable for expression in prokaryotes (AAR47151)
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
CC terminal sequence of IL-2 receptor gamma chain, and are used to
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
CC are used to obtain the protein given in AAR47151.
XX
SQ Sequence 347 AA;

```

```

Query Match 100.0%; Score 34; DB 15; Length 347;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 LQNLVIP 7
DB 126 lqnlvip 132

```

```

RESULT 6
ID AAR47148 standard; Protein; 369 AA.
XX
AC AAR47148;
XX

```

```

DT 13-JUN-1994 (first entry)
XX
DE IL-2 receptor gamma chain.
XX

```

```

KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KW rheumatoid arthritis; transplant rejection; primer; PCR;
KW polymerase chain reaction; amplification; ss.
XX
OS Homo sapiens.
XX

```

```

FH Key Location/Qualifiers
FT Peptide 1..22

```

FT /label= Sig_peptide
 XX EP578932-A.
 PN 19-JAN-1994.
 XX
 PD
 XX
 PF 22-APR-1993; 93EP-0106561.
 XX
 PR 23-APR-1992; 92JP-0104947.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX (SUGA/) SUGAMURA K.
 PI Aaso H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
 PI Suzuki M, Takeshita T;
 XX WPI: 1994-017546/03.
 DR N-PSDB: AAQ54828.
 XX
 PT DNA and protein sequences of IL-2 gamma chain - useful as immune
 PT regulatory agents for treatment of e.g. rheumatoid arthritis and
 PT transplant rejection
 XX
 PS Disclosure; Page 16-17, 29-30; 50pp; English.
 XX
 CC The human IL-2 receptor gamma chain preform (AAK47148), including the
 CC signal peptide, is encoded by the sequence given in AAQ54828. The
 CC mature protein (AAK47149) is encoded by sequence AAQ54829. A soluble
 CC form of IL-2 receptor gamma chain (AAK47150) is encoded by AAQ54830,
 CC while a soluble form suitable for expression in prokaryotes (AAK47151)
 CC is encoded by AAQ54831. Primers 1-6 (AAQ54820-25) are based on the N-
 CC terminal sequence of IL-2 receptor gamma chain, and are used to
 CC isolate IL2 receptor gamma chain receptor cDNA. Primers AAQ54826-27
 CC are used to obtain the protein given in AAK47151.
 XX
 SQ Sequence 369 AA;
 QY 1 LQNLVIP 7
 Db 148 lqnlvip 154
 Query Match 100.0%; Score 34; DB 15; Length 369;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 7
 AAK59094
 ID AAK59094 standard; Protein: 369 AA.
 XX
 AC AAK59094;
 XX
 DT 04-MAY-1995 (first entry)
 XX
 DE Murine IL-2R gamma.
 XX
 KW Murine IL2-R gamma; X-linked severe combined immunodeficiency;
 KW XSCID; Interleukin.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT Domain /note= "signal peptide"
 FT /note= "transmembrane domain"
 FT MISC-difference 331
 FT Modified-site /note= "Corresponding codon CAG"
 FT /label= N-glycosylation_site
 FT Modified-site 75..77
 FT /label= N-glycosylation_site

FT Modified-site 84..86
 FT /label= N-glycosylation_site
 FT Modified-site 96..98
 FT /label= N-glycosylation_site
 FT Modified-site 159..161
 FT /label= N-glycosylation_site
 FT Modified-site 255..257
 FT /label= N-glycosylation_site
 XX
 PN WO9420641-A.
 XX
 PD 15-SEP-1994.
 XX
 PF 10-MAR-1994; 94WO-US02891.
 XX
 PR 12-MAR-1993; 93US-0031143.
 PR 14-SEP-1993; 93US-0121435.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Leonard WJ, McBride WO, Noguchi M;
 PI N-PSDB: AAQ71977.
 DR WPI: 1994-303046/37.
 DR N-PSDB: AAQ71977.
 XX
 PT Diagnosis of X-linked severe combined immunodeficiency (XSCID) -
 PT comprises detecting mutated IL-2R gamma gene, also vectors and
 PT transgenic animals containing the mutated gene
 XX
 PS Example 1; Fig 7; 98pp; English.
 XX
 CC AAK71977 is the DNA sequence of murine IL-2R gamma AAK59094,
 CC this was used in the development of a claimed method for the
 CC diagnosis of X-linked severe combined immunodeficiency (XSCID),
 CC in female carriers and male sufferers.
 XX
 SQ Sequence 369 AA;
 QY 1 LQNLVIP 7
 Db 148 lqnlvip 154
 Query Match 100.0%; Score 34; DB 15; Length 369;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 8
 AAK31646
 ID AAK31646 standard; Protein: 482 AA.
 XX
 AC AAK31646;
 XX
 DT 21-MAY-1998 (first entry)
 XX
 DE Human cytokine receptor gc chain-Ig fusion protein.
 XX
 KW Cytokine receptor; gamma common chain; gc chain; human;
 KW blocking agent; monoclonal antibody; CP.88; Immunological disease;
 KW myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
 KW insulin-dependent diabetes; inflammatory bowel disease;
 KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;
 KW graft versus host disease; psoriasis; immunosuppressive; therapy.
 XX
 OS Chimeric - Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..254
 FT /note= "gc chain N-terminal region"
 FT Protein 255..482
 FT /note= "IgG1 constant region"
 FT Region 255..264

FT FT /note= "IgG1 hinge region"
 FT Domain 264...482
 FT /note= "IgG1 CH2 and CH3 constant domains!"
 XX
 PN WO9743416-A1.
 XX
 XX 20-NOV-1997.
 PD
 XX
 PF 09-MAY-1997; 97WO-US07870.
 XX
 PR 10-MAY-1996; 96US-0017466.
 XX
 PA (BIOJ) BIOGEN INC.
 XX
 PI Benjamin CD, Burkly LC, Hession C, Whitty A;
 DR WPI: 1998-008885/01.
 DR N-PSDB: AAT97439.
 XX
 PT Blocking agents of the gamma common chain of cytokine receptors -
 PT particularly monoclonal antibodies, used to induce T cell anergy for
 PT treatment of immunological diseases
 XX
 PS Example 1; Page 79-80; 11pp; English.
 XX
 CC This polypeptide comprises a fusion between the N-terminal 254
 CC amino acids of the human mature cytokine receptor gamma common (gc)
 CC chain and the hinge region and CH2 and CH3 constant domains of
 CC human IgG1. The fusion was expressed from clone pLB001 (see
 CC AAT97439) in COS-7 cells, and used to generate murine anti-human gc
 CC specific monoclonal antibodies (Mabs), including CP.B8 produced by
 CC hybridoma ATCC HB 12107. The invention provides compositions and
 CC methods for inhibiting cytokine signalling using gc chain blocking
 CC agents for the treatment of immunological diseases such as
 CC myasthenia gravis, rheumatoid arthritis, lupus, multiple sclerosis,
 CC insulin-dependent diabetes, inflammatory bowel disease, sympathetic
 CC ophthalmia, uveitis, allergy, asthma, parasitic infection, graft
 CC vs. host disease or psoriasis. A preferred gc blocking agent is
 CC Mab CP.B8 or its Fab fragment (see also AAW31647-48).
 XX
 SQ Sequence 482 AA:
 OY 1 LQNLVIP 7
 DB 148 lqnlvip 154
 Query Match 100.0%; Score 34; DB 19; Length 482;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 9
 AAY92202
 ID AAY92202 standard; Protein; 691 AA.
 XX
 AC AAY92202;
 XX
 DT 01-AUG-2000 (first entry)
 XX
 DE Fusion polypeptide 603, IL-4 trap.
 XX
 KW IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;
 KW cytostatic; immunomodulator; osteopathic.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN WO200018932-A2.
 XX
 PD 06-APR-2000.
 XX
 PF 22-SEP-1999; 99WO-US22045.

XX
 PR 25-SEP-1998; 98US-0101858.
 PR 19-MAY-1999; 99US-0313942.
 XX
 PA (REGG-) REGENERON PHARM INC.
 XX
 PI Stahl N, Yancopoulos GD;
 DR WPI: 2000-293165/25.
 DR N-PSDB: AAA09044.
 XX
 PT Isolated nucleic acid molecule for treating cytokine-related diseases
 PT or disorders encodes a fusion polypeptide capable of binding a cytokine
 PT to form a nonfunctional complex
 XX
 PS Example 6; Fig 22A-D; 152pp; English.
 XX
 CC This sequence shows fusion polypeptide 603, which is capable of
 CC binding cytokine IL-4 to form a non-functional complex.
 CC The invention concerns production of antagonists to any cytokine that
 CC utilizes an alpha specificity determining component, which when combined
 CC with the cytokine, binds to a first beta signal transducing component to
 CC form a non-functional intermediate which then binds to a second beta
 CC signal transducing component causing beta-receptor dimerization, the
 CC soluble alpha specificity determining component of the receptor
 CC (sr-alpha) and the extracellular domain of the first beta signal
 CC transducing component of the cytokine receptor (beta-1) are combined to
 CC form heterodimers (sr-alpha:beta-1) that act as antagonist to the
 CC cytokine by binding the cytokine to form a non-functional complex. The
 CC receptor components are shared by cytokines such as the CNTF (ciliary
 CC neurotrophic factor) family of cytokines. The invention provides the
 CC basis for the development of IL-6 antagonists, as they show that if, in
 CC the presence of a ligand, a non-functional intermediate complex,
 CC consisting of the ligand, its alpha receptor and its beta-1 receptor
 CC component, can be formed, it will effectively block the action of the
 CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers
 CC of the extracellular domains of the alpha specificity determining
 CC components of their receptors and the extracellular domain of gp130.
 CC The resultant heterodimers, function as high-affinity traps, rendering
 CC the cytokine inaccessible to form a signal transducing complex with the
 CC native membrane-bound forms of their receptor. The nucleic acids and
 CC polypeptides are useful for treating cytokine-related diseases or
 CC disorders such as osteoporosis and primary and secondary effects of
 CC cancer including multiple myeloma or cachexia.
 XX
 SQ Sequence 691 AA:
 OY 1 LQNLVIP 7
 DB 148 lqnlvip 154
 Query Match 100.0%; Score 34; DB 21; Length 691;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 10
 AAY92201
 ID AAY92201 standard; Protein; 694 AA.
 XX
 AC AAY92201;
 XX
 DT 01-AUG-2000 (first entry)
 XX
 DE Fusion polypeptide 424, IL-4 trap.
 XX
 KW IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;
 KW cytostatic; immunomodulator; osteopathic.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX

PN WO200018932-A2.
 XX 06-APR-2000.
 PD
 XX
 XX 22-SEP-1999; 99WO-US22045.
 PF
 XX 25-SEP-1998; 98US-0101858.
 PR 19-MAY-1999; 99US-0313942.
 XX
 XX (REG-) REGENERON PHARM INC.
 PA
 XX Stahl N, Yancopoulos GD;
 PI
 XX WPI; 2000-293165/25.
 DR N-PSDB; AAA09043.
 DR
 XX
 XX Isolated nucleic acid molecule for treating cytokine-related diseases
 PT or disorders encodes a fusion polypeptide capable of binding a cytokine
 PT to form a nonfunctional complex
 PS
 XX Example 6; Fig 21; 152pp; English.
 PS
 CC This sequence shows fusion polypeptide 424, which is capable of
 CC binding cytokine IL-4 to form a non-functional complex.
 CC The invention concerns production of antagonists to any cytokine that
 CC utilizes an alpha specificity determining component, which when combined
 CC with the cytokine, binds to a first beta signal transducing component to
 CC form a non-functional intermediate which then binds to a second beta
 CC signal transducing component causing beta-receptor dimerization, the
 CC soluble alpha specificity determining component of the receptor
 CC (SR-alpha) and the extracellular domain of the first beta signal
 CC transducing component of the cytokine receptor (beta-1) are combined to
 CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the
 CC cytokine by binding the cytokine to form a non-functional complex. The
 CC receptor components are shared by cytokines such as the CNTF (ciliary
 CC neurotrophic factor) family of cytokines. The invention provides the
 CC basis for the development of IL-6 antagonists, as they show that if, in
 CC the presence of a ligand, a non-functional intermediate complex,
 CC consisting of the ligand, its alpha receptor and its beta-1 receptor
 CC component, can be formed, it will effectively block the action of the
 CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers
 CC of the extracellular domains of the alpha specificity determining
 CC components of their receptors and the extracellular domain of gp130.
 CC The resultant heterodimers, function as high-affinity traps, rendering
 CC the cytokine inaccessible to form a signal transducing complex with the
 CC native membrane-bound forms of their receptor. The nucleic acids and
 CC polypeptides are useful for treating cytokine-related diseases or
 CC disorders such as osteoporosis and primary and secondary effects of
 CC cancer including multiple myeloma or cachexia.
 CC
 XX
 SO Sequence 694 AA;

Query Match 100.0%; Score 34; DB 21; Length 694;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
 Db 148 lqnlvip 154
 |||||
 |||||

RESULT 11
 ID AAY92203 standard; Protein: 694 AA.
 XX
 XX AAY92203;
 AC
 XX
 XX 01-AUG-2000 (first entry)
 DT
 XX Fusion polypeptide 622, IL-4 trap.
 DE
 XX IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;
 KM

KW cyostatic; immunomodulator; osteopathic.
 XX
 XX Synthetic.
 OS
 OS Homo sapiens.
 XX
 XX WO200018932-A2.
 PN
 XX 06-APR-2000.
 PD
 XX
 XX 22-SEP-1999; 99WO-US22045.
 PF
 XX 25-SEP-1998; 98US-0101858.
 PR 19-MAY-1999; 99US-0313942.
 XX
 XX (REG-) REGENERON PHARM INC.
 PA
 XX Stahl N, Yancopoulos GD;
 PI
 XX WPI; 2000-293165/25.
 DR N-PSDB; AAA09045.
 DR
 XX
 XX Isolated nucleic acid molecule for treating cytokine-related diseases
 PT or disorders encodes a fusion polypeptide capable of binding a cytokine
 PT to form a nonfunctional complex
 PS
 XX Example 6; Fig 23A-D; 152pp; English.
 PS
 CC This sequence shows fusion polypeptide 622, which is capable of
 CC binding cytokine IL-4 to form a non-functional complex.
 CC The invention concerns production of antagonists to any cytokine that
 CC utilizes an alpha specificity determining component, which when combined
 CC with the cytokine, binds to a first beta signal transducing component to
 CC form a non-functional intermediate which then binds to a second beta
 CC signal transducing component causing beta-receptor dimerization, the
 CC soluble alpha specificity determining component of the receptor
 CC (SR-alpha) and the extracellular domain of the first beta signal
 CC transducing component of the cytokine receptor (beta-1) are combined to
 CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the
 CC cytokine by binding the cytokine to form a non-functional complex. The
 CC receptor components are shared by cytokines such as the CNTF (ciliary
 CC neurotrophic factor) family of cytokines. The invention provides the
 CC basis for the development of IL-6 antagonists, as they show that if, in
 CC the presence of a ligand, a non-functional intermediate complex,
 CC consisting of the ligand, its alpha receptor and its beta-1 receptor
 CC component, can be formed, it will effectively block the action of the
 CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers
 CC of the extracellular domains of the alpha specificity determining
 CC components of their receptors and the extracellular domain of gp130.
 CC The resultant heterodimers, function as high-affinity traps, rendering
 CC the cytokine inaccessible to form a signal transducing complex with the
 CC native membrane-bound forms of their receptor. The nucleic acids and
 CC polypeptides are useful for treating cytokine-related diseases or
 CC disorders such as osteoporosis and primary and secondary effects of
 CC cancer including multiple myeloma or cachexia.
 CC
 XX
 SO Sequence 694 AA;

Query Match 100.0%; Score 34; DB 21; Length 694;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
 Db 148 lqnlvip 154
 |||||
 |||||

RESULT 12
 ID AAW72225 standard; Protein: 1374 AA.
 XX
 XX AAW72225;
 AC
 XX

DT 13-JAN-1999 (first entry)
XX
DE HSV-2 strain SB5 Contig ID 15 ORF#20b protein.
XX
XX HSV-2 strain SB5: immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor.
XX
OS Herpes simplex virus type 2.
XX
XX WO9820016-A1.
XX
PD 14-MAY-1998.
XX
XX 31-OCT-1997; 97WO-US20016.
XX
XX 09-JUN-1997; 97US-0049018.
PR 04-NOV-1996; 96US-0030279.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;
PI Esser KM, Leary JY;
XX
DR WPI: 1998-286847/25.
DR N-PSDB; AAV62176.
XX
XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal
XX
PS Claim 10; Page 141-142; 748bp; English.
XX
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.
CC Based on homology, this sequence is a major capsid protein.
CC The proteins can be used for the treatment or prevention of disease, to
CC induce an immunological response in a mammal or to identify inhibitors,
CC activators or novel antivirals. Antagonists of the proteins can be used
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC it can also be used to induce an immunological response in a mammal.
XX
SQ Sequence 1374 AA;

QY 1 LONLVIP 7
| | | : | : |
DB 839 lqnmvvp 845

RESULT 13
AAW69753
ID AAW69753 standard; protein; 1374 AA.
XX
AC AAW69753;
XX
DT 02-NOV-1998 (first entry)
XX
DE Herpes simplex virus-1 VP5 protein.
XX
XX Herpes simplex virus-1; HSV-1; UL15; VP5; UL19; antiviral agent;
KW identification.
XX
OS Herpes simplex virus type 1.
XX
XX EP860700-A2.
XX
PD 26-AUG-1998.
XX

PF 16-FEB-1998; 98EP-0301123.
XX
XX 21-FEB-1997; 97US-0038914.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Del Vecchio A;
XX
DR WPI: 1998-439489/38.
XX
PT Screening assays for antiviral agents - based on inhibition of
PT herpes simplex UL15 and VP5 interaction or function
XX
XX Claim 10; Page 11-16; 18pp; English.
XX
XX Methods have been developed for: (1) identifying an antiviral agent
CC useful in treating infection with herpes viruses, comprising: (a)
CC providing UL15 of herpes simplex virus-1 (HSV-1) or a functional
CC derivative or homologue thereof; (b) permitting UL15 or its derivative
CC or homologue to come into association with a test compound; (c)
CC providing VP5 of HSV-1 or a functional derivative or homologue thereof;
CC (d) screening for inhibition of the interaction between UL15 or its
CC derivative or homologue and VP5 or its derivative or homologue and VP5
CC or its derivative or homologue, where such inhibition is indicative that
CC the test compound is an antiviral agent; (2) identifying an antiviral
CC agent useful in treating infection with herpes viruses, comprising: (a)
CC providing VP5 of herpes simplex virus-1 (HSV-1) or a functional
CC derivative or homologue thereof; (b) permitting VP5 or its derivative
CC or homologue to come into association with a test compound; (c)
CC providing UL15 of HSV-1 or a functional derivative or homologue
CC thereof; (d) screening for inhibition of the interaction between the
CC VP5 or its derivative or homologue and UL15 or its derivative or
CC homologue, where such inhibition is indicative that the test compound
CC is an antiviral agent. The present sequence represents the HSV-1 VP5
CC (or UL15) protein.
XX
SQ Sequence 1374 AA;

QY 1 LONLVIP 7
| | | : | : |
DB 839 lqnmvvp 845

RESULT 14
AAW72224
ID AAW72224 standard; protein; 1384 AA.
XX
AC AAW72224;
XX
DT 13-JAN-1999 (first entry)
XX
DE HSV-2 strain SB5 Contig ID 15 ORF#20a protein.
XX
XX HSV-2 strain SB5; immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor.
XX
OS Herpes simplex virus type 2.
XX
XX WO9820016-A1.
XX
PD 14-MAY-1998.
XX
XX 31-OCT-1997; 97WO-US20016.
XX
XX 09-JUN-1997; 97US-0049018.
PR 04-NOV-1996; 96US-0030279.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX

XX Chan JY, Debrowski-Amara CE, Delvecchio AM, Dillon SB;
PI Esser KM, Leary JY;
XX WPI: 1998-286847/25.
DR N-PSDB: AAV62176.
XX
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal
XX
PS Claim 10: Page 140-141: 748bp; English.
XX
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.
CC Based on homology, this sequence is a major capsid protein.
CC The proteins can be used for the treatment or prevention of disease, to
CC induce an immunological response in a mammal or to identify inhibitors,
CC activators or novel antivirals. Antagonists of the proteins can be used
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC it can also be used to induce an immunological response in a mammal.
XX
SQ Sequence 1384 AA;

Query Match 91.28; Score 31; DB 19; Length 1384;
Best Local Similarity 71.4%; Pred. NO. 5.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 1 LONLVIP 7
|||:|:|
Db 849 lqnmvvp 855

RESULT 15
AAW72117
ID AAW72117 standard; Protein: 1396 AA.
XX
AC AAW72117;
XX
DT 18-DEC-1998 (first entry)
XX
DE HSV-2 strain SB5 Contig ID 15 ORF#4 protein.
XX
KM HSV-2 strain SB5; immunological response induction; therapy;
KM antiviral identification; viral protein inhibitor.
XX
OS Herpes simplex virus type 2.
XX
PN W09820016-A1.
XX
PD 14-MAY-1998.
XX
PF 31-OCT-1997; 97WO-US20016.
XX
PR 09-JUN-1997; 97US-0049018.
PR 04-NOV-1996; 96US-0030279.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Chan JY, Debrowski-Amara CE, Delvecchio AM, Dillon SB;
PI Esser KM, Leary JY;
XX
DR WPI: 1998-286847/25.
DR N-PSDB: AAV62159.
XX
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal
XX
PS Claim 10: Page 88-89: 748bp; English.
XX

CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.
CC Based on homology, this sequence is a major capsid protein.
CC The proteins can be used for the treatment or prevention of disease, to
CC induce an immunological response in a mammal or to identify inhibitors,
CC activators or novel antivirals. Antagonists of the proteins can be used
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC it can also be used to induce an immunological response in a mammal.
XX
SQ Sequence 1396 AA;

Query Match 91.28; Score 31; DB 19; Length 1396;
Best Local Similarity 71.4%; Pred. NO. 5.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 1 LONLVIP 7
|||:|:|
Db 861 lqnmvvp 867

Search completed: August 20, 2002, 11:07:17
Job time: 6102 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:10:36 ; Search time 100.84 Seconds

(without alignments)
6.670 Million cell updates/sec

Title: US-09-824-286-15

Perfect score: 34

Sequence: 1 LQNMVIP 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : PIR-71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	369	2	149280
2	34	100.0	369	2	A42565
3	34	100.0	373	2	A55718
4	33	97.1	496	2	E70142
5	31	91.2	216	2	AE2548
6	31	91.2	1374	1	VCBE17
7	30	88.2	267	2	S62488
8	30	88.2	303	2	S32583
9	30	88.2	346	2	T14049
10	30	88.2	361	1	QOCVPT
11	30	88.2	513	2	E97393
12	30	88.2	513	2	AG2611
13	30	88.2	553	2	S53080
14	30	88.2	689	2	T19999
15	30	88.2	1028	2	B24785
16	29	85.3	171	2	T23696
17	29	85.3	184	2	T51655
18	29	85.3	217	2	T14621
19	29	85.3	246	2	E89884
20	29	85.3	261	2	T46138
21	29	85.3	279	2	AC2165
22	29	85.3	345	2	CE4489
23	29	85.3	346	2	T17105
24	29	85.3	346	2	T13835
25	29	85.3	346	2	T14056
26	29	85.3	379	2	S55845
27	29	85.3	513	2	T27355
28	29	85.3	535	2	S44827
29	29	85.3	537	2	T04745

ALIGNMENTS

30	29	85.3	622	2	S17402	parasporal crystal
31	29	85.3	652	2	G95177	hypothetical prote
32	29	85.3	652	2	S47979	alb protein precu
33	29	85.3	652	2	E98044	hypothetical prote
34	29	85.3	661	2	T08314	hypothetical prote
35	29	85.3	662	2	H84984	bo-type ubiquinol
36	29	85.3	662	2	H71676	DNA gyrase chain B
37	29	85.3	662	2	E97738	2-amin-4-hydroxy-
38	29	85.3	744	2	A82822	NADH-ubiquinone ox
39	28	82.4	174	2	G83661	myb-related trans
40	28	82.4	188	2	T51676	ABC transporter At
41	28	82.4	216	2	B82020	cell division Arp
42	28	82.4	216	2	AC2397	ATP-binding protei
43	28	82.4	243	2	G71507	hypothetical prote
44	28	82.4	243	2	G71507	glycerol-3-phospha
45	28	82.4	307	2	AG2017	

RESULT 1

149280
Interleukin-2 receptor gamma chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Feb-1997 #sequence.revision 27-Feb-1997 #text.change 20-Jun-2000
C:Accession: I49280; A47514; JN0592; JN0775; S37582; I53398
R:Caio, X.; Kozak, C.A.; Liu, Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 8464-8468, 1993
A:Title: Characterization of cDNAs encoding the murine interleukin 2 receptor (IL-2R)
A:Reference number: A47514; MUID:93391374
A:Accession: I49280
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369 <CAO>
A:Cross-References: EMBL:U21795; NID:g727349; PIDN:AAA64279.1; PID:g727350
A:Accession: A47514
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-369 <RE2>
A:Cross-References: GB:L20048; NID:g404067; PIDN:AAA9286.1; PID:g404068
R:Kumaki, S.; Kondo, M.; Takeshita, T.; Asao, H.; Nakamura, M.; Sugamura, K.
Biochem. Biophys. Res. Commun. 193, 356-363, 1993
A:Title: Cloning of the mouse interleukin 2 receptor gamma chain: Demonstration of fu
A:Reference number: JN0592; MUID:93277575
A:Accession: JN0592
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-369 <KUM>
A:Cross-References: DDBJ:D13565; NID:g303684; PIDN:BAA02760.1; PID:g303685
R:Kobayashi, N.; Nakagawa, S.; Minami, Y.; Taniguchi, T.; Kono, T.
Gene 130, 303-304, 1993
A:Title: Cloning and sequencing of the cDNA encoding a mouse IL-2 receptor gamma.
A:Reference number: JN0775; MUID:93366191
A:Accession: JN0775
A:Molecule type: mRNA
A:Residues: 1-369 <KOB>
A:Cross-References: GB:D13821; NID:g436045; PIDN:BAA02974.1; PID:g436046
R:Chiu, R.K.; Dougherty, G.J.
submitted to the EMBL Data Library, October 1993
A:Description: Regulation of CD44-mediated cellular adhesion by the IL-2 R gamma cha
A:Reference number: S37582
A:Accession: S37582
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-350, S', 352-366, S', 368-369 <CHI>
A:Cross-References: EMBL:X75337
R:Disanto, J.P.; Certain, S.; Wilson, A.; MacDonald, H.R.; Avner, P.; Fischer, A.; de
Eur. J. Immunol. 24, 3014-3018, 1994
A:Title: The murine interleukin-2 receptor gamma chain gene: organization, chromosoma
A:Reference number: I53398; MUID:95104285
A:Accession: I53398
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-369 <RES>
 A:Cross-references: GB:S75852; NID:9861554; PIDN:AA32904.1; PID:9861555
 C:Genetics:
 A:Gene: IL-2Rgamma
 A:Introns: 39/1; 90/2; 152/1; 199/3; 254/1; 286/2; 308/3
 C:Complex: The high affinity receptor is a heterotrimer of alpha (see PIR:UHMS2), beta
 eptors.
 C:Function:
 A:Description: receptor for interleukin-2
 A:Pathway: Interleukin-2 stimulated growth and differentiation of T cells, B cells, NK
 C:Superfamily: Interleukin-2 receptor gamma chain
 C:Keywords: cytokine receptor; glycoprotein; transmembrane protein
 F:1-22/Domain: signal sequence #status predicted <SIC>
 F:23-369/Product: interleukin-2 receptor gamma chain #status predicted <MAT>
 F:256-284/Domain: transmembrane #status predicted <TM>
 F:1/1;5,84,96,159,164,306/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 34; DB 2; Length 369;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
 Db 148 LQNLVIP 154

RESULT 2
 A42565
 Interleukin-2 receptor gamma chain - human
 C:Species: Homo sapiens (man)
 C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
 C:Accession: A42565; A46591; F54332
 R:Takeishi, T.; Aseo, H.; Ohnani, K.; Ishii, N.; Kumaki, S.; Tanaka, N.; Munakata, H.;
 Science 257, 379-382, 1992
 A:Title: Cloning of the gamma chain of the human IL-2 receptor.
 A:Reference number: A42565; MWID:92335883
 A:Accession: A42565
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid; protein
 A:Residues: 1-369 <RAK>
 A:Cross-references: GB:D11086; NID:q303611; PIDN:BA01857.1; PID:q319890
 A:Experimental source: MOLT beta lymphoid cells
 A:Note: sequence extracted from NCBI backbone (NCBIP:109167)
 R:Moguchi, M.; Adelstein, S.; Cao, X.; Leonard, W.J.
 J. Biol. Chem. 268, 13601-13608, 1993
 A:Title: Characterization of the human interleukin-2 receptor gamma chain gene.
 A:Reference number: A46591; MWID:93293887
 A:Accession: A46591
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-369 <RES>
 A:Cross-references: GB:L12183; NID:q307056; PIDN:AA59145.1; PID:q307058
 R:Puck, J.M.; Deschenes, S.M.; Porter, J.C.; Dutra, A.S.; Brown, C.J.; Willard, H.F.; He
 Hum. Mol. Genet. 2, 1099-1104, 1993
 A:Title: The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-link
 A:Reference number: F54332; MWID:94004847
 A:Accession: F54332
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-369 <RE2>
 A:Cross-references: GB:L19546; NID:q349631; PIDN:AA37524.1; PID:q349632
 C:Genetics:
 A:Gene: GDB:IL2RG; SCIDX1; IMD4
 A:Cross-references: GDB:134807; OMIM:308380
 A:Map position: Xq13.1-Xq13.1
 A:Introns: 39/1; 90/2; 152/1; 198/3; 253/1; 285/2; 308/3
 A:Note: defects are associated with an X-linked form of severe combined immunodeficiency
 C:Superfamily: Interleukin-2 receptor gamma chain
 C:Keywords: cytokine receptor; duplication; immunodeficiency; severe combined immunodeficiency

Query Match 100.0%; Score 34; DB 2; Length 369;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
 Db 148 LQNLVIP 154

RESULT 3
 A55718
 Interleukin-2 receptor gamma chain precursor - dog
 C:Species: Canis lupus familiaris (dog)
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 23-Jul-1999
 C:Accession: A55718
 R:Henlton, P.S.; Somberg, R.L.; Flimant, V.M.; Puck, J.M.; Patterson, D.F.; Felsburg
 Genomics 23, 69-74, 1994
 A:Title: IL-2Rgamma gene microdeletion demonstrates that canine X-linked severe combi
 A:Reference number: A55718; MWID:9510114
 A:Accession: A55718
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-373 <HEN>
 A:Cross-references: GB:U04361; NID:q517411; PIDN:AA48403.1; PID:q517412
 C:Superfamily: Interleukin-2 receptor gamma chain
 C:Keywords: cytokine receptor; duplication

Query Match 100.0%; Score 34; DB 2; Length 373;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
 Db 148 LQNLVIP 154

RESULT 4
 E70142
 Glu-tRNA amidotransferase, subunit A (gluA) homolog - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 22-Jun-1999
 C:Accession: E70142
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
 son, D.; Peterson, J.; Kervase, A.R.; Queckebush, J.; Salzberg, S.; Hanson, M.; Vu
 Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MWID:98065943
 A:Accession: E70142
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-496 <RLS>
 A:Cross-references: GB:AE001140; GB:AE000783; NID:q2688233; PIDN:AA66715.1; PID:q268
 A:Experimental source: strain B31
 C:Superfamily: indoleacetamide hydrolase

Query Match 97.1%; Score 33; DB 2; Length 496;
 Best Local Similarity 85.7%; Pred. No. 12;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
 Db 395 LQNLVIP 401

RESULT 5
 AE2348
 hypothetical protein a117672 [imported] - Anabaena sp. (strain PCC 7120) plasmid PCC7
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AE2548
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iritani,
N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium *Ana*
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2548
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-216 <KUR>
A:Cross-references: GB:AP003602; PIDN:BAW7315.1; PID:g17134757; GSPDB:GN00181
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all7672
A:Genome: plasmid

Query Match 91.2%; Score 31; DB 2; Length 216;
Best Local Similarity 71.4%; Pred. No. 14;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7
|||||
Db 64 LONLVIP 70

RESULT 6
VCBE1
major capsid protein - human herpesvirus 1 (strain 17)
C:Species: human herpesvirus 1
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 16-Jun-2000
C:Accession: A27239; A30084
R:Davidson, A.J.; Scott, J.E.
J. Gen. Virol. 67, 2279-2286, 1986
A:Title: DNA sequence of the major capsid protein gene of herpes simplex virus type 1.
A:Reference number: A27239; MUID:87010565
A:Accession: A27239
A:Molecule type: DNA
A:Residues: 1-1374 <DNA>
A:Cross-references: GB:X04467; NID:g59857; PIDN:CAA28154.1; PID:g59859
R:McGoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perz
J. Gen. Virol. 69, 1531-1574, 1988
A:Title: The complete DNA sequence of the long unique region in the genome of herpes sim
A:Reference number: A30083; MUID:88274327
A:Accession: A30084
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1374 <MCG>
A:Cross-references: GB:X14112; NID:g1944536; PIDN:CAA32332.1; PID:g59519; GB:DD00317
C:Genetics:
A:Gene: UL19
C:Superfamily: varicella-zoster virus major capsid protein
C:Keywords: capsid protein

Query Match 91.2%; Score 31; DB 1; Length 1374;
Best Local Similarity 71.4%; Pred. No. 11e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7
|||||
Db 839 LONLVIP 845

RESULT 7
S62488
hypothetical protein SPAC4G8.11c - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C>Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 31-Jan-2000
C:Accession: T38856; S62488
R:Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, October 1995

A:Reference number: Z21745
A:Accession: T38856
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-267 <BA2>
A:Cross-references: EMBL:Z56276; NID:g1022345; PIDN:CAA91212.1; PID:g1022356; GSPDB:(
A:Experimental source: strain 972h-; cosmid c4G8
C:Genetics:
A:Gene: SPAC4G8.11c
A:Map position: 1L
A:Introns: 114/3; 203/3

Query Match 88.2%; Score 30; DB 2; Length 267;
Best Local Similarity 85.7%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7
|||||
Db 6 LONLVIP 12

RESULT 8
S32583
structural protein VP40 - Marburg virus (strain Popp)
C:Species: Marburg virus
A:Variety: strain Popp
C>Date: 10-Dec-1993 #sequence_revision 08-Sep-1995 #text_change 08-Oct-1999
C:Accession: S32583; S44051; S28569
R:Bukreyev, A.A.; Volchkov, V.E.; Bilnov, V.M.; Netesov, S.V.
FEBS Lett. 322, 41-46, 1993
A:Title: The VP3s and VP40 proteins of filoviruses. Homology between Marburg and Ebol
A:Reference number: S32582; MUID:93245956
A:Accession: S32583
A:Molecule type: mRNA
A:Residues: 1-303 <BUK>
A:Cross-references: EMBL:X64406; NID:g60631; PIDN:CAA45749.1; PID:g60633
A:Experimental source: strain Popp
R:Bukreyev, A.A.; Volchkov, V.E.; Bilnov, V.M.; Dryga, S.A.; Netesov, S.V.
submitted to the EMBL Data Library, January 1994
A:Description: Full-length nucleotide sequence of Marburg virus Popp strain: The comp
A:Reference number: S44049
A:Accession: S44051
A:Molecule type: genomic RNA
A:Residues: 1-303 <BUV>
A:Cross-references: EMBL:Z29337; NID:g450908; PIDN:CAA82538.1; PID:g450911
A:Experimental source: strain Popp

Query Match 88.2%; Score 30; DB 2; Length 303;
Best Local Similarity 71.4%; Pred. No. 36;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7
|||||
Db 146 LONLVIP 152

RESULT 9
T14049
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - *Lamna nasus* mitochondrion (fra
C:Species: mitochondrion *Lamna nasus*
A:Note: porbeagle shark
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-Jan-2000
C:Accession: T14049
R:Naylor, G.J.P.; Martin, A.P.; Mattison, E.G.; Brown, W.M.
submitted to the EMBL Data Library, February 1997
A:Description: Interrelationships of Lamniform Sharks: Testing Phylogenetic Hypotheses
A:Reference number: Z17808
A:Accession: T14049
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-346 <NAV>

A:Cross-references: EMBL:U91427; NID:g2257811; PID:g2257812; PIDN:AAB63136.1
A:Genetics:
A:Genome: mitochondrion
A>Note: NADH2
A:Superfamily: NADH dehydrogenase (ubiquinone) chain 2
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 88.2%; Score 30; DB 2; Length 346;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 QNLVIP 7
|||||
Db 273 QNLVIP 278

RESULT 10

OCVPT

ALI protein - potato yellow mosaic virus (isolate Venezuela)

C:Species: potato yellow mosaic virus

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jun-2000

C:Accession: J00364

R:Cutts, R.H.A.; Coffin, R.S.; Roberts, E.J.F.; Hamilton, W.D.O.

J. Gen. Virol. 72, 1515-1520, 1991

A:Title: The nucleotide sequence of the infectious cloned DNA components of potato yellow

A:Reference number: J00362; MID:91311403

A:Accession: J00364

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1361 <CON>

A:Cross-references: GB:D00940; NID:g222458; PIDN:BAA00782.1; PID:g222459

C:Genetics:

A:Map position: segment A

C:Superfamily: tomato golden mosaic virus ALI protein

Query Match 88.2%; Score 30; DB 1; Length 361;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 QNLVIP 7
|||||
Db 32 QNLVIP 38

RESULT 11

E97393

hypothetical protein AGR_C_491 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C:Species: Agrobacterium tumefaciens

C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002

C:Accession: E97393

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194

A:Accession: E97393

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-513 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK86102.1; PID:g1515183; GSPDB:GN00169

C:Genetics:

A:Gene: AGR_C_491

A:Map position: circular chromosome

Query Match 88.2%; Score 30; DB 2; Length 513;
Best Local Similarity 85.7%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 QNLVIP 7
|||||

Db 187 QNLVIP 193

RESULT 12

AG2611

conserved hypothetical protein Atu0287 [imported] - Agrobacterium tumefaciens (strain

C:Species: Agrobacterium tumefaciens

C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C:Accession: AG2611

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McCl

er, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AG2611

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-513 <KUR>

A:Cross-references: GB:AE008688; PIDN:AL41309.1; PID:g17738619; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu0287

A:Map position: circular chromosome

Query Match 88.2%; Score 30; DB 2; Length 513;
Best Local Similarity 85.7%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 QNLVIP 7
|||||
Db 187 QNLVIP 193

RESULT 13

S53080

hypothetical protein YMR258c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YMR920.12c

C:Species: Saccharomyces cerevisiae

C>Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 04-Mar-2000

C:Accession: S53080

R:Hunt, S.; Bowman, S.

submitted to the EMBL Data Library, March 1995

A:Reference number: S53069

A:Accession: S53080

A:Molecule type: DNA

A:Residues: 1-553 <HUN>

A:Cross-references: EMBL:Z48639; NID:g732924; PID:g732936; GSPDB:GN00013; MIPS:YMR258

C:Genetics:

A:Gene: MIPS:YMR258c

A:Map position: 13R

C:Superfamily: Saccharomyces cerevisiae hypothetical protein YMR258c

Query Match 88.2%; Score 30; DB 2; Length 553;
Best Local Similarity 71.4%; Pred. No. 71;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 QNLVIP 7
|||||
Db 192 QNLVIP 198

RESULT 14

T19999

hypothetical protein C47D12.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T19999

R:Gajadaty, S.

submitted to the EMBL Data Library, March 1996

A:Reference number: Z19209

A:Accession: T19999

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-689 <WILL>

A:Cross-references: EMBL:Z69902; PIDN:CAA93767.1; GSPDB:GN00020; CESP:C47D12.2

A:Experimental source: clone C47D12

C:Genetics:

A:Gene: CESP:C47D12.2

A:Map position: 2

A:Introns: 48/2; 97/2; 167/1; 196/2; 277/1; 439/3; 478/3; 542/3

Query Match

88.2%; Score 30; DB 2; Length 689;

Best Local Similarity 71.4%; Pred. No. 91;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNLVIP 7

Db 392 QNLVIP 398

RESULT 15

B24785

hypothetical protein 1028 - slime mold (Dictyostelium discoideum) transposon DIRS-1 (fr

C:Species: Dictyostelium discoideum

C>Date: 23-Aug-1987 #sequence_revision 23-Aug-1987 #text_change 18-Jun-1993

C:Accession: B24785

R:Capello, J.; Handelsman, K.; Lodish, H.F.

Cell 43, 105-115, 1985

A:Title: Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted ter

A:Reference number: A94654; MUID:86079481

A:Accession: B24785

A:Molecule type: DNA

A:Residues: 1-1028 <CAP>

Query Match

88.2%; Score 30; DB 2; Length 1028;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QNLVIP 7

Db 447 QNLVIP 452

Search completed: August 20, 2002, 11:10:37
Job time: 5667 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:33:01 ; Search time 55.29 Seconds

(Without alignments)
4.902 Million cell updates/sec

Title: US-09-824-286-15

Perfect score: 34

Sequence: 1 LQNLVTP 7

Scoring table: BLOSUM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	369	1 CYRG_HUMAN	P31785 homo sapien
2	34	100.0	369	1 CYRG_MOUSE	P34902 mus musculu
3	34	100.0	373	1 CYRG_CANFA	P40321 canis famli
4	33	97.1	496	1 GATV_BORBU	O51317 borrelia bu
5	31	91.2	1374	1 VCAP_HSV11	P06491 herpes simp
6	30	88.2	267	1 YADB_SCHPO	O09836 schizosacch
7	30	88.2	283	1 VP40_MABVM	P35260 marburg vir
8	30	88.2	303	1 VP40_MABVM	O03040 marburg vir
9	30	88.2	361	1 VAL1_PYMVP	P27258 potato yell
10	30	85.3	553	1 YMBA_YEAST	O04847 saccharomyc
11	29	85.3	379	1 CYRG_BOVIN	O95118 bos taurus
12	29	85.3	379	1 YLEU_DEBOC	P48012 debaryomyce
13	29	85.3	535	1 YMA2_CAEEL	P34447 caenorhabdi
14	29	85.3	622	1 ALIB_STRPN	O45743 bacillus th
15	29	85.3	652	1 CYOB_BUCAI	O51933 streptococc
16	29	85.3	662	1 CYOB_BUCAI	P57543 buchera ap
17	28	82.4	325	1 V653_STRPY	O9407 streptococc
18	28	82.4	342	1 HRCA_LERIN	O51867 leptospira
19	28	82.4	359	1 RFL_CHLMU	O9P116 chlamydia m
20	28	82.4	359	1 RFL_CHLMU	O84026 chlamydia m
21	28	82.4	441	1 YXER_BACSU	P34950 bacillus su
22	28	82.4	663	1 CYOB_ECOLI	P18401 escherichia
23	28	82.4	666	1 COL4_BRAJA	P98057 bradyrhizob
24	28	82.4	732	1 ADD1_CAEEL	O90900 caenorhabdi
25	28	82.4	1228	1 ECM_HUMAN	O13201 homo sapien
26	28	82.4	1517	1 YD22_SCHPO	O10250 schizosacch
27	28	82.4	1597	1 RRL1_YEAST	P53552 saccharomyc
28	28	82.4	1950	1 UBRI_YEAST	P19812 saccharomyc
29	27	79.4	221	1 V700_RICPR	O9cmm4 rickettsia
30	27	79.4	268	1 NIKE_ECOLI	P33594 escherichia
31	27	79.4	275	1 NIFH_METWP	O50218 methanococc
32	27	79.4	296	1 VP23_HSV6U	P24436 human herpe
33	27	79.4	405	1 S3AE_BACSU	P49782 bacillus su

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	369 AA.
CYRG_HUMAN				
ID	CYRG_HUMAN			
AC	P31785:			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cytokine receptor common gamma chain precursor (Gamma-C) (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64) (CD132 antigen).			
GN	IL2RG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92335883; PubMed=1631559;			
RA	Takeshita T., Asao H., Ohtani K., Ishii N., Kumaki S., Tanaka N.,			
RT	Munakata H., Nakamura M., Sugamura K.;			
RL	"Cloning of the gamma chain of the human IL-2 receptor.";			
RN	Science 257:379-382(1992).			
RP	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RC	MEDLINE=93293887; PubMed=8514792;			
RA	Noguchi M., Adelstein S., Cao X., Leonard W.J.;			
RT	"Characterization of the human Interleukin-2 receptor gamma chain			
RL	gene.";			
RL	J. Biol. Chem. 268:13601-13608(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND VARIANTS ASP-114 AND ASN-153.			
RX	MEDLINE=94004847; PubMed=8401490;			
RA	Puck J.M., Deschenes S.M., Porter J.C., Dutra A.S., Brown C.J.,			
RT	Willard H., Henthorn P.S.;			
RL	"The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated			
RN	in X-linked severe combined immunodeficiency, SCIDX1.";			
RL	Hum. Mol. Genet. 2:1099-1104(1993).			
RP	[4]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RX	MEDLINE=94090315; PubMed=8266076;			
RA	Kondo M., Takeshita T., Ishii N., Nakamura M., Watanabe S.,			
RT	Arai K.-I., Sugamura K.;			
RL	"Sharing of the interleukin-2 (IL-2) receptor gamma chain between			
RN	receptors for IL-2 and IL-4.";			
RL	Science 262:1874-1877(1993).			
RN	[5]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RX	MEDLINE=94090317; PubMed=8266078;			
RA	Russell S.M., Kkegan A.D., Harada N., Nakamura Y., Noguchi M.,			
RT	Leonard P., Friedmann M.C., Miyajima A., Puri R.K., Paul W.E.,			
RL	Leonard W.J.;			
RT	"Interleukin-2 receptor gamma chain: a functional component of the			
RL	interleukin-4 receptor.";			
RN	Science 262:1880-1883(1993).			
RP	[6]			
RP	IDENTIFICATION AS A IL-7R SUBUNIT.			

34	27	79.4	428	1 AROA_CAMJE	P52312 campylobact
35	27	79.4	517	1 MB13_YEAST	O92247 saccharomyc
36	27	79.4	563	1 YCUA_YEAST	P25355 saccharomyc
37	27	79.4	583	1 NTR_BEYVE	P38500 betula verr
38	27	79.4	591	1 ARSF_HUMAN	P54799 homo sapien
39	27	79.4	594	1 NTR_SPIOL	P05314 spinacia ol
40	27	79.4	612	1 Y146_RICPR	O92461 rickettsia
41	27	79.4	716	1 HGFL_MOUSE	P26928 mus musculu
42	27	79.4	726	1 BFL1_HUMAN	P50747 h biotin--p
43	27	79.4	795	1 SYER_BUCAI	P57230 buchera ap
44	27	79.4	878	1 YBM5_SCHPO	O10337 schizosacch
45	27	79.4	890	1 YOUN_ECOLI	P39838 escherichia

RX MEDLINE-94090316; PubMed-8266077;
 RA Noguchi M., Nakamura Y., Russell S.M., Ziegler S.F., Tsang M., Cao X.,
 RA Leonard W.J.;
 RT "Interleukin-2 receptor gamma chain: a functional component of the
 RT interleukin-7 receptor.";
 RL Science 262:1877-1880(1993).
 RN [17]
 RP 3D-STRUCTURE MODELING OF 57-248.
 RX MEDLINE-95111955; PubMed-7529123;
 RA Babrough P., Hedgecock C.J., Richards W.G.;
 RT "The interleukin-2 and interleukin-4 receptors studied by molecular
 RT modelling";
 RL Structure 2:839-851(1994).
 RN [18]
 RP VARIANTS XSCID PHE-115; CYS-240 AND IL2-241.
 RX MEDLINE-94130970; PubMed-8299698;
 RA Disanto J.P., Dautry-Varsat A., Certain S., Fischer A.,
 RA de Saint Basile G.;
 RT "Interleukin-2 (IL-2) receptor gamma chain mutations in X-linked
 RT severe combined immunodeficiency disease result in the loss of
 RT high-affinity IL-2 receptor binding.";
 RL Eur. J. Immunol. 24:475-479(1994).
 RN [19]
 RP VARIANT XSCID LYS-68.
 RX MEDLINE-94375038; PubMed-8088810;
 RA Markiewicz S., Subtil A., Dautry-Varsat A., Fischer A.,
 RA de Saint Basile G.;
 RT "Detection of three nonsense mutations and one missense mutation in
 RT the interleukin-2 receptor gamma chain gene in SCIDx1 that
 RT differentially affect the mRNA processing.";
 RL Genomics 21:291-293(1994).
 RN [10]
 RP VARIANT XSCID HIS-162.
 RX MEDLINE-94300093; PubMed-8027558;
 RA Ishii N., Asao H., Kimura Y., Takeshita T., Nakamura M., Tsuchiya S.,
 RA Kono T., Maeda M., Uchiyama T., Sugamura K.;
 RT "Impairment of ligand binding and growth signaling of mutant IL-2
 RT receptor gamma-chains in patients with X-linked severe combined
 RT immunodeficiency.";
 RL J. Immunol. 153:1310-1317(1994).
 RN [11]
 RP VARIANT XSCID ASN-39.
 RX MEDLINE-95023933; PubMed-7937790;
 RA Disanto J.P., Rieux-Laucat F., Dautry-Varsat A., Fischer A.,
 RA de Saint Basile G.;
 RT "Defective human interleukin 2 receptor gamma chain in an atypical X
 RT chromosome-linked severe combined immunodeficiency with peripheral T
 RT cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9466-9470(1994).
 RN [12]
 RP VARIANTS XSCID CYS-226 AND HIS-226.
 RX MEDLINE-95397841; PubMed-7668284;
 RA Pepper A.E., Buckley R.H., Small T.N., Puck J.M.;
 RT "Two mutational hotspots in the interleukin-2 receptor gamma chain
 RT gene causing human X-linked severe combined immunodeficiency.";
 RL Am. J. Hum. Genet. 57:564-571(1995).
 RN [13]
 RP VARIANT XSCID SSR-183.
 RX MEDLINE-96013903; PubMed-7557965;
 RA Clark P.A., Lester T., Genet S., Jones A.M., Hendriks R.,
 RA Levinsky R.L., Kinon C.;
 RT "Screening for mutations causing X-linked severe combined
 RT immunodeficiency in the IL-2R gamma chain gene by single-strand
 RT conformation polymorphism analysis.";
 RL Hum. Genet. 96:427-432(1995).
 RN [14]
 RP VARIANT XSCID GLN-237 G-H-M INS.
 RX MEDLINE-95164726; PubMed-7860773;
 RA Puck J.M., Pepper A.E., Bedard P.-M., Laframboise R.;
 RT "Female germ line mosaicism as the origin of a unique IL-2 receptor
 RT gamma-chain mutation causing X-linked severe combined
 RT immunodeficiency.";
 RL J. Clin. Invest. 95:895-899(1995).

RN [15]
 RP VARIANT XSCID GLN-293.
 RX MEDLINE-95190013; PubMed-7883965;
 RA Schmalstieg F.C., Leonard W.J., Noguchi M., Berg M., Rudloff H.E.,
 RA Denney R.M., Dave S.K., Brooks E.G., Goldman A.S.;
 RT "Missense mutation in exon 7 of the common gamma chain gene causes a
 RT moderate form of X-linked combined immunodeficiency.";
 RL J. Clin. Invest. 95:1169-1173(1995).
 RN [16]
 RP VARIANT XSCID ARG-115.
 RX MEDLINE-97042245; PubMed-8900089;
 RA Stephan V., Mann V., Le Delst F., Dirksen U., Broeker B.,
 RA Mueller-Fleckenstein I., Horneff G., Schroten H., Fischer A.,
 RA de Saint Basile G.;
 RT "Atypical X-linked severe combined immunodeficiency due to possible
 RT spontaneous reversion of the genetic defect in T cells.";
 RL New Engl. J. Med. 335:1563-1567(1996).
 RN [17]
 RP VARIANT XSCID GLN-285.
 RX MEDLINE-97295088; PubMed-9150740;
 RA Jones A.M., Clark P.A., Katz F., Genet S., McMahon C., Alterman L.,
 RA Cant A., Kinon C.;
 RT "B-cell-negative severe combined immunodeficiency associated with a
 RT common gamma chain mutation.";
 RL Hum. Genet. 99:677-680(1997).
 RN [18]
 RP VARIANT XSCID CYS-222.
 RX MEDLINE-98064061; PubMed-9399950;
 RA Sharfe N., Shahar M., Roifman C.M.;
 RT "An interleukin-2 receptor gamma chain mutation with normal thymus
 RT morphology.";
 RL J. Clin. Invest. 100:3036-3043(1997).
 RN [19]
 RP FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
 CC INTERLEUKINS.
 CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
 CC -1- PROBABLY ALSO THE IL-13 RECEPTORS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A SEVERE COMBINED
 CC IMMUNODEFICIENCY, WHICH IS KNOWN AS AGAMMAGLOBULINEMIA, SWISS TYPE
 CC OR X-LINKED SEVERE COMBINED IMMUNODEFICIENCY DISEASE (XSCID).
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD132 entry;
 CC NAME="http://www.ncbi.nlm.nih.gov/prov/cd/cd132.htm".
 CC -1- DATABASE: NAME=NCBI; NOTE=X-linked SCID mutation database;
 CC WWW="http://www.nhgri.nih.gov/DIR/GMNB/SCID/".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-ch).
 CC -----
 DR EMBL; D11086; BAA01857.1; -;
 DR EMBL; L12183; AAA59145.1; JOINED.
 DR EMBL; L12178; AAA59145.1; JOINED.
 DR EMBL; L12176; AAA59145.1; JOINED.
 DR EMBL; L12177; AAA59145.1; JOINED.
 DR EMBL; L12179; AAA59145.1; JOINED.
 DR EMBL; L12180; AAA59145.1; JOINED.
 DR EMBL; L12181; AAA59145.1; JOINED.
 DR EMBL; L12182; AAA59145.1; JOINED.
 DR EMBL; L19546; AAC37524.1; -;
 DR PIR; A42565; A42565.
 DR PDB; 1IIM; 26-JAN-95.
 DR PDB; 1ILN; 26-JAN-95.
 DR MIM; 308380; -;
 DR MIM; 300400; -;
 DR InterPro: IPR002996; CRIA.
 DR InterPro: IPR003961; FN.III.
 DR InterPro: IPR003531; Hematopo_receptor_S.F1.

```

Query Match          100.0%; Score 34; DB 1; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7
DB 148 LONLVIP 154

RESULT 2
CYRG_MOUSE          STANDARD; PRT; 369 AA.
ID CYRG_MOUSE
AC P34902;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cytokine receptor common gamma chain precursor (Gamma-C)
DE (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).
GN IL2RG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93377575; PubMed-8503926;
RA Kumaki S., Kondo M., Takeshita T., Asao H., Nakamura M., Sugamura K.;
RT "Cloning of the mouse interleukin 2 receptor gamma chain:
RT demonstration of functional differences between the mouse and human
RT receptors.";
RL Biochem. Biophys. Res. Commun. 193:356-363(1993).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN-CBA/CA.
RC MEDLINE-93391374; PubMed-8378320;
RA Cao X., Kozak C.A., Liu Y.J., Noguchi M., O'Connell E., Leonard W.J.;
RT "Characterization of cDNAs encoding the murine interleukin 2 receptor
RT (IL-2R) gamma chain: chromosomal mapping and tissue specificity of
RT IL-2R gamma chain expression.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:8464-8468(1993).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE-9336191; PubMed-8359699;
RA Kobayashi N., Nakagawa S., Minami Y., Taniguchi T., Kono T.;
RT "Cloning and sequencing of the cDNA encoding a mouse IL-2 receptor
RT gamma.";
RL gene 130:303-304(1993).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE-95104285; PubMed-7805729;
RA Disanto J.P., Certain S., Wilson A., Macdonald H.R., Avner P.;
RA Fischer A., de Saint Basile G.;
RT "The murine interleukin-2 receptor gamma chain gene: organization,
RT chromosomal localization and expression in the adult thymus.";
RL Eur. J. Immunol. 24:3014-3018(1994).
[5]
RP SEQUENCE FROM N.A.
RX STRAIN-B6.S;
RC MEDLINE-96341745; PubMed-8750189;
RA Chiu R.K., Droll A., Cooper D.L., Dougherty S.T., Dirks J.F.,
RA Dougherty G.J.;
RT "Molecular mechanisms regulating the hyaluronan binding activity of
RT the adhesion protein CD44.";
RL J. Neurooncol. 26:231-239(1995).
[6]
RP FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
RP INTERLEUKINS.
CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D13821; BAA02974.1; -
DR EMBL: U21795; AAA64279.1; -
DR EMBL: D13565; BAA02760.1; -
DR EMBL: L20048; AAA39286.1; -
DR EMBL: S75852; AAB32904.1; -
DR EMBL: S75844; AAB32904.1; JOINED.
DR EMBL: S75845; AAB32904.1; JOINED.
DR EMBL: S75847; AAB32904.1; JOINED.
DR EMBL: S75848; AAB32904.1; JOINED.
DR EMBL: S75849; AAB32904.1; JOINED.
DR EMBL: S75850; AAB32904.1; JOINED.
DR EMBL: S75851; AAB32904.1; JOINED.
DR EMBL: X75337; CAA53085.1; -
DR PIR: JN0592; JN0592.
DR PIR: JN0775; JN0775.
DR HSSP: P31785; ILLN.
DR MGD: MGI:96551; IL2rg.
DR InterPro: IPR002996; CRLA.
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003531; Hematopo_receptor_S_F1.
DR Pfam: PF00041; fn3; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS01355; HEMATOPO_REC_S_F1; 1.
DR Receptor: Transmembrane; Glycoprotein; Signal.
KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT CHAIN 1 22
FT SIGNAL 1 22
FT DOMAIN 23 369
FT TRANSMEM 264 284
FT DOMAIN 285 369
FT DOMAIN 151 250
FT DISULFID 62 72
FT DISULFID 102 115
FT CARBOHYD 71 71
FT CARBOHYD 75 75
FT CARBOHYD 77 75
FT CARBOHYD 84 84
FT CARBOHYD 96 96
FT CARBOHYD 159 159
FT CARBOHYD 164 164
SQ SEQUENCE 369 AA; 42241 MW; CB2D5AB459077AC7 CRC64;

Query Match          100.0%; Score 34; DB 1; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7
DB 148 LONLVIP 154

RESULT 3
CYRG_CANFA          STANDARD; PRT; 373 AA.
ID CYRG_CANFA
AC P40321;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cytokine receptor common gamma chain precursor (Gamma-C)
DE (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).
GN IL2RG.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
[1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE-Spleen;
RX MEDLINE=95130114; PubMed=7829104;
RA Henthorn P.S., Somberg R.L., Flamlant V.M., Puck J.M., Patterson D.F.,
RA Felsburg P.J.;
RT "IL-2R gamma gene microdeletion demonstrates that canine x-linked
RT severe combined immunodeficiency is a homologue of the human
RT disease.";
RL Genomics 23:69-74(1994).
CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
CC INTERLEUKINS.
CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC PROBABLY ALSO THE IL-13 RECEPTORS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A CANINE X-LINKED
CC SEVERE COMBINED IMMUNODEFICIENCY.
CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U04361; AAC48403.1; -
DR HSSP: P31785; IILN.
DR InterPro: IPR002996; CRIA.
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003531; Hematopo_receptor_S_F1.
DR Pfam: PF00041; FN3; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS01355; HEMATOPO_REC_S_F1; 1.
KM Receptor: Transmembrane; Glycoprotein; signal.
FT SIGNAL 1 22
FT CHAIN 1 373
FT DOMAIN 23 261
FT TRANSMEM 262 283
FT TRANSMEM 284 373
FT DOMAIN 151 249
FT DISULFID 62 72
FT DISULFID 102 115
FT CARBOHYD 24 24
FT CARBOHYD 71 71
FT CARBOHYD 75 75
FT CARBOHYD 84 84
FT CARBOHYD 159 159
FT CARBOHYD 164 164
FT CARBOHYD 249 249
SQ SEQUENCE 373 AA; 42516 MW; 03A0DE1F8B089D8B CRC64;

Query Match 100.0%; Score 34; DB 1; Length 373;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
Db 148 LQNLVIP 154

RESULT 4
GATA_BORBU
ID GATA_BORBU STANDARD; PRT; 496 AA.
AC 051317;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamy-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.-) (Glu-ADT
DE subunit A).
GN GATA OR BB0342.

```

```

OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_Taxid=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerecavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Ueterbach T., Watney L., McDonald L., Arlrich P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
CC -1- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
CC GLN-TRNA(GLN) THROUGH THE TRANSAMIDATION OF MISACYLATED GLU-
CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE -> ADP
CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.
CC -1- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001140; AAC6715.1; -
DR TIGR: BB0342; -
DR InterPro: IPR000120; Amidase.
DR Pfam: PF01425; Amidase; 1.
DR PROSITE: PS00571; AMIDASES; 1.
KM Protein biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 496 AA; 55544 MW; D375DEE2DD042710 CRC64;

Query Match 97.1%; Score 33; DB 1; Length 496;
Best Local Similarity 85.7%; Pred. No. 5.4;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
Db 395 LQNLVIP 401

RESULT 5
VCAP_HSV11
ID VCAP_HSV11 STANDARD; PRT; 1374 AA.
AC P06491;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major capsid protein (MCP) (capsid protein VP3).
DE UL19.
GN Herpes simplex virus (type 1 / strain 17).
OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_Taxid=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of

```

```

RT herpes simplex virus type 1."
RL J. Gen. Virol. 69:1531-1574(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE-87010565; PubMed-3020164;
RT Davison B.A.J., Scott J.E.;
RT "DNA sequence of the major capsid protein gene of herpes simplex
RL virus type 1."
RN J. Gen. Virol. 67:2279-2286(1986).
RN [3]
RP SEQUENCE OF 202-211 AND 607-616.
RA MEDLINE-93019027; PubMed-1328483;
RT Davison M.D., Rixon F.J., Davison A.J.;
RT "Identification of genes encoding two capsid proteins (VP24 and VP26)
RT of herpes simplex virus type 1."
RL J. Gen. Virol. 73:2709-2713(1992).
CC -1- FUNCTION: MAJOR PROTEIN OF THE ICOSAHEDRAL CAPSID.
CC -1- SUBUNIT: MAIN COMPONENT OF THE HEXAVALENT, AND PROBABLY THE
CC PENTAVALENT CAPSOMERES.
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES MAJOR CAPSID PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10879; BAA01665.1; -
DR EMBL: X14112; CAA32332.1; -
DR EMBL: X04467; CAA28154.1; -
DR PIR: A27239; VCBRI7.
DR PIR: A30084; A30084.
DR InterPro: IPR000912; Herpes_MCP.
DR Pfam: PF03122; Herpes_MCP.1.
DR PRINTS: PR00235; HSVCAPSIDMCP.
KM Coat protein.
SQ SEQUENCE 1374 AA; 149083 MW; 859C76E2EADDE05B7 CRC64;

Query Match
Best Local Similarity 91.2%; Score 31; DB 1; Length 1374;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
DB 839 LQNLVIP 845

RESULT 6
YADB_SCHPO STANDARD; PRT; 267 AA.
AC 009836;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DE Hypothetical 31.4 kDa protein C4G8.11C in chromosome I.
GN SPAC4G8.11C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Badcock K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: SOME, TO YEAST ATP10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z56276; CAA91212.1; -
KM Hypothetical protein.
SQ SEQUENCE 267 AA; 31381 MW; B38390F48D4D835C CRC64;

Query Match
Best Local Similarity 88.2%; Score 30; DB 1; Length 267;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
DB 6 LQNLVIP 12

RESULT 7
VP40_MABVM STANDARD; PRT; 283 AA.
ID VP40_MABVM
AC P35260;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Matrix protein VP40.
GN VP40.
OS Marburg virus (strain Musoke).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC Filovirus.
OX NCBI_TaxID=33727;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92327834; PubMed-1626422;
RA Feldmann H., Muehberger E., Randolph A., Will C., Killey M.P.,
RA Sanchez A., Klenk H.D.;
RT "Marburg virus, a filovirus: messenger RNAs, gene order, and
RT regulatory elements of the replication cycle."
RL Virus Res. 24:1-19(1992).
RN [2]
RP REVISIONS.
RA Feldmann H.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z12132; CAA78116.1; ALT_TERM.
KM Matrix protein.
SQ SEQUENCE 283 AA; 31653 MW; C338D448B2345265 CRC64;

Query Match
Best Local Similarity 88.2%; Score 30; DB 1; Length 283;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
DB 146 LQNLVIP 152

RESULT 8
VP40_MABVP STANDARD; PRT; 303 AA.
ID VP40_MABVP
AC 003040;

```

```

DT 01-FEB-1994 (Rel. 28, Last Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Matrix protein VP40 (VP3).
GN VP40.
OS Marburg virus (strain Popp).
OS VIRUSES: ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC Filoviridae.
OC NCBI_TaxID=33728;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=93245956; PubMed=8482365;
RA Bukreyev A.A., Volchkov V.E., Bilinov V.M., Netesov S.V.;
RT "The VP35 and VP40 proteins of filoviruses. Homology between Marburg
and Ebola viruses.";
RL FEBS Lett. 322:41-46(1993).
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=96028047; PubMed=7487490;
RA Bukreyev A.A., Volchkov V.E., Bilinov V.M., Dryga S.A., Netesov S.V.;
RT "The complete nucleotide sequence of the Popp (1967) strain of Marburg
virus: a comparison with the Musoke (1980) strain.";
RL Arch. Virol. 140:1589-1600(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64406; CAA45749.1; -
DR EMBL; Z29337; CAA82538.1; -
DR PIR; S32583; S32583.
DR PIR; S28569; S28569.
DR PIR; S44051; S44051.
KM Matrix protein.
SQ SEQUENCE 303 AA; 33734 MW; 4AB3B45402C961DD CRC64;

Query Match 88.2%; Score 30; DB 1; Length 303;
Best Local Similarity 71.4%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
Db 146 LQNLVIP 152

RESULT 9
VAL1_PYMV STANDARD; PRT; 361 AA.
AC P27258;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE AL1 protein.
OS Potato yellow mosaic virus (isolate Venezuela).
OC VIRUSES: ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=10828;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=91311403; PubMed=1856690;
RA Coutts R.H.A., Coffin R.S., Roberts E.J.F., Hamilton W.D.O.;
RT "The nucleotide sequence of the infectious cloned DNA components of
RT potato yellow mosaic virus.";
RL J. Gen. Virol. 72:1515-1520(1991).
CC -----
CC -1- SIMILARITY: BELONGS TO GEMINIVIRUSES AL1 PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00940; BAA00782.1; -
DR PIR; J00364; OQCVPT.
DR InterPro; IPR001191; Gemini1_AL1.
DR Pfam; PF00799; Gemini1_AL1.1.
DR PRINTS; PR00227; GEMCOAT1.1.
DR Prodom; PD000736; Gemini1_AL1.1.
KW ATP-binding.
FT NP_BIND 222 229 ATP (POTENTIAL).
SQ SEQUENCE 361 AA; 40850 MW; 5627A33BF1264383 CRC64;

Query Match 88.2%; Score 30; DB 1; Length 361;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
Db 32 LQNLVIP 38

RESULT 10
YMB8_YEAST STANDARD; PRT; 553 AA.
ID YMB8_YEAST
AC 004847;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 64.4 kDa protein in PER111-TIF11 intergenic region.
GN YMR258C OR YM9920.12C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RX SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48639; CAA88585.1; -
DR SGD; S0004871; YMR258C.
KM Hypothetical protein.
SQ SEQUENCE 553 AA; 64405 MW; 2F764964A6C23EC2 CRC64;

Query Match 88.2%; Score 30; DB 1; Length 553;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
Db 192 LQNLVIP 198

RESULT 11
CYRG_BOVIN STANDARD; PRT; 379 AA.
ID CYRG_BOVIN
AC 095118;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```



```

DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cytokine receptor common gamma chain precursor (Gamma-C)
DE (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).
GN IL2RG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96268473; PubMed=8672241;
RA Yoo J., Stone R.T., Solinas-Toldo S., Fries R., Beattie C.W.;
RT "Cloning and chromosomal mapping of bovine interleukin-2 receptor
gamma gene."
CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
CC INTERLEUKINS.
CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC PROBABLY ALSO THE IL-13 RECEPTORS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U33748; AAB07812.1; -.
DR HSSP: P31785; IILN.
DR InterPro: IPR002996; CRIA.
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003531; Hematopo_receptor_S_F1.
DR Pfam: PF00041; FN3; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS01355; HEMATOPO_REC_S_F1; 1.
KW Receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 1 379 CYTOKINE RECEPTOR COMMON GAMMA CHAIN.
FT DOMAIN 23 269 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 270 290 POTENTIAL.
FT DOMAIN 291 379 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 158 256 FIBRONECTIN TYPE-III.
FT DISULFID 68 78 POTENTIAL.
FT DISULFID 109 122 POTENTIAL.
FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 379 AA; 43037 MW; 33CFAD9C9B032178 CRC64;

Query Match 85.3%; Score 29; DB 1; Length 379;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

DE Hypothetical 43.6 kDa protein.
OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=27300;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 26077 / R91;
RX MEDLINE=95321019; PubMed=7597851;
RA Iserentant D., Verachtert H.;
RT "Cloning and sequencing of the LEU2 homologue gene of Schwanniomyces
occidentalis."
CC Yeast 11:467-473(1995).
CC -1- SIMILARITY: TO YEAST BUD3.
CC -1- CAUTION: IS INDICATED BY REF.1 TO BE A 3-ISOPROPYLALATE
CC DEHYDROGENASE (EC 1.1.1.85) (LEU2), BUT THIS IS MOST PROBABLY A
CC MISASSIGNMENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X79823; CAA56224.1; -.
DR InterPro: IPR000219; RhogGEF.
DR SMART: SM00325; RhogGEF; 1.
KW Hypothetical protein.
SQ SEQUENCE 379 AA; 43586 MW; 6C846CD0B53B0BC CRC64;

```

```

Query Match 85.3%; Score 29; DB 1; Length 379;
Best Local Similarity 71.4%; Pred. No. 35;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LONLVIP 7
Db 46 LONLVIP 52

RESULT 13
YMA2_CAEEL STANDARD; PRT; 535 AA.
ID YMA2_CAEEL
AC P34447;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 56.5 kDa protein F54P2.2 in chromosome III.
GN F54P2.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Spoat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans."
RL Nature 368:32-38(1994).

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L23645; AAA28046.1; -
DR PIR: S44827; S44827.
DR Morph: P54F2.2; CE00195.
DR Hypothetical protein.
KM SEQUENCE 535 AA: 56492 MW: 731E133D327AD9A CRC64;
SQ
Query Match      85.3%; Score 29; DB 1; Length 535;
Best Local Similarity 85.7%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 LQNLVIP 7
Db 150 LQNLVIP 156
RESULT 14
C2AC_BACTU STANDARD; PRT; 622 AA.
ID C2AC_BACTU
AC Q45743;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 40, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Pesticidal crystal protein cry2Ac (Insecticidal delta-endotoxin
DE CryIIA(C)) (Crystalline entomocidal protoxin) (70 kDa crystal protein).
GN CRY2AC OR CRYIIA(C) OR CRYIIC.
OS Bacillus thuringiensis.
OG plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SHANGHAI 1 / S-1;
RX MEDLINE-91340086; PubMed-1651878;
RT Wu D., Cao X.L., Bai Y.Y., Aronson A.I.;
RT "Sequence of an operon containing a novel delta-endotoxin gene from
RT Bacillus thuringiensis.";
RL FEMS Microbiol. Lett. 65:31-36(1991).
CC -1- FUNCTION: PROMOTES COLLOIDSMORPHIC LYSIS BY BINDING TO THE MIDGUT
CC EPITHELIAL CELLS OF LEPIDOPTERAN LARVAE. HAS LOW ACTIVITY ON
CC DIPTERAN LARVAE.
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57252; CAA40536.1; -
DR InterPro: IPR001178; Endotoxin.
DR Pfam: PF00555; endotoxin.1.
KM Toxin; Sporulation; Plasmid.
SQ SEQUENCE 622 AA: 69729 MW: E0B5FAD37BF8299 CRC64;

```

```

Query Match      85.3%; Score 29; DB 1; Length 622;
Best Local Similarity 71.4%; Pred. No. 58;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LQNLVIP 7
Db 77 LQNLVIP 83
RESULT 15
ALIB_STRPN STANDARD; PRT; 652 AA.
ID ALIB_STRPN
AC Q51933;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oligopeptide-binding protein alib precursor.
GN ALIB OR SP1527.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R800;
RX MEDLINE-94328326; PubMed-8051706;
RA Allouin G., de Philip P., Claverys J.-P.;
RA "Three highly homologous membrane-bound lipoproteins participate in
RT oligopeptide transport by the Aml system of the Gram-positive
RT Streptococcus pneumoniae.";
RL J. Mol. Biol. 241:44-58(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGRA;
RX MEDLINE-21357209; PubMed-11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfe E., Khouli H., Wolf A.M., Ullrich T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiboul S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC FOR OLIGOPEPTIDES. PROBABLY AN OLIGOPEPTIDE BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z16082; CAA78896.1; -
DR EMBL: AE007448; AAK75616.1; -
DR TIGR: SP1527;
DR InterPro: IPR000914; SBP_bac-5.
DR Pfam: PF00496; SBP_bac-5; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE: PS01040; SBP_BACTERIAL_5; 1.
KM Peptide transport; Transport; Membrane; Lipoprotein; Signal;
KW Complete proteome.
FT SIGNAL 1 24 PROBABLE.
FT CHAIN 25 652 OLIGOPEPTIDE-BINDING PROTEIN ALIB.
FT LIPID 25 25 N-ACYL DIGLYCERIDE (PROBABLE).

```

FT CONFLICT 55 55 A -> R (IN REF. 1).
FT CONFLICT 79 80 SL -> HI (IN REF. 1).
FT CONFLICT 123 124 LO -> FE (IN REF. 1).
FT CONFLICT 501 501 G -> E (IN REF. 1).
SQ SEQUENCE 652 AA; 72562 MM; 169B67FD78CF0CF CRC64;

Query Match 85.3%; Score 29; DB 1; Length 652;
Best Local Similarity 71.4%; Pred. No. 61;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7
1:111:1
DB 363 LRNLVVP 369

Search completed: August 20, 2002, 11:33:02
Job time: 1451 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:31:55 ; Search time 191.14 seconds
(without alignments)
6.335 Million cell updates/sec

Title: US-09-824-286-15
Perfect score: 34
Sequence: 1 LONLVIP 7

Scoring table: BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp-organellar:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp_bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	34	100.0	273	2	Q93DQ0	Q93DQ0 uncultured
2	34	100.0	273	2	Q93DP7	Q93DP7 uncultured
3	32	94.1	3380	5	Q90943	Q90943 Locusta mig
4	31	91.2	358	13	Q9W616	Q9W616 brachydanio
5	31	91.2	358	13	Q9PW55	Q9PW55 brachydanio
6	31	91.2	1374	12	P89442	P89442 herpes simp
7	30	88.2	225	12	Q9WHP8	Q9WHP8 sinaleoa tom
8	30	88.2	235	12	Q93178	Q93178 geminivirid
9	30	88.2	346	8	Q21187	Q21187 lamna nasus
10	30	88.2	362	13	Q9W617	Q9W617 brachydanio
11	30	88.2	372	12	Q9PYV2	Q9PYV2 xestia c-ni
12	30	88.2	373	11	Q9CXB2	Q9CXB2 mus musculus
13	30	88.2	634	5	Q44116	Q44116 drosophila
14	30	88.2	634	5	Q9V978	Q9V978 drosophila
15	30	88.2	689	5	Q18668	Q18668 caenorhabdi
16	30	88.2	1007	5	Q96848	Q96848 dictyostell

17	30	88.2	1007	5	Q23897	Q23897 dictyostell
18	29	85.3	171	5	Q94184	Q94184 caenorhabdi
19	29	85.3	184	10	Q9XRE2	Q9XRE2 arabidopsi
20	29	85.3	217	10	Q40292	Q40292 beta vulgar
21	29	85.3	246	16	Q990W7	Q990W7 staphylococ
22	29	85.3	261	10	Q9SMW1	Q9SMW1 arabidopsi
23	29	85.3	345	10	Q9SK24	Q9SK24 arabidopsi
24	29	85.3	346	8	Q21195	Q21195 megachasma
25	29	85.3	346	8	Q21676	Q21676 alopias pel
26	29	85.3	346	8	Q21129	Q21129 alopias pel
27	29	85.3	461	10	Q9SDP4	Q9SDP4 allium cepa
28	29	85.3	513	5	Q9XW90	Q9XW90 caenorhabdi
29	29	85.3	537	10	Q9S2V0	Q9S2V0 arabidopsi
30	29	85.3	633	4	Q96C78	Q96C78 homo sapien
31	29	85.3	641	4	Q94927	Q94927 homo sapien
32	29	85.3	661	17	Q52001	Q52001 halobacteri
33	29	85.3	662	16	Q9ZDU7	Q9ZDU7 rickettsia
34	29	85.3	662	16	Q921W1	Q921W1 rickettsia
35	29	85.3	744	16	Q9PG19	Q9PG19 xylella fas
36	29	85.3	839	5	Q10652	Q10652 caenorhabdi
37	29	85.3	980	5	Q9H6G1	Q9H6G1 halocynthia
38	28	82.4	79	2	Q99223	Q99223 lactococcus
39	28	82.4	99	15	Q999N3	Q999N3 human immun
40	28	82.4	151	17	Q972F9	Q972F9 sulfolobus
41	28	82.4	174	16	Q9KGG7	Q9KGG7 bacillus ha
42	28	82.4	188	10	Q9ZTD0	Q9ZTD0 arabidopsi
43	28	82.4	216	2	P95356	P95356 neisseria g
44	28	82.4	216	16	Q9KIR3	Q9KIR3 neisseria m
45	28	82.4	216	16	Q9JW11	Q9JW11 neisseria m

ALIGNMENTS

RESULT 1	
Q93DQ0	PRELIMINARY; PRT; 273 AA.
AC Q93DQ0;	
DT 01-DEC-2001 (TREMUREL. 19, Created)	
DT 01-DEC-2001 (TREMUREL. 19, Last sequence update)	
DE 2,5-DIKETO-D-GLUCONIC ACID REDUCTASE.	
OS uncultured bacterium.	
OC Bacteria; environmental samples.	
OX NCBI_TaxID=77133;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=21417047; PubMed=11526025;	
RA Eschenfeldt W.H., Stols L., Rosenbaum H., Khambatla Z.S.,	
RA Quate-Randall E., Wu S., Kilgore D.C., Trent J.D., Donnelly M.I.;	
RT "Dna from uncultured organisms as a source of 2,5-diketo-D-gluconic	
RT acid reductases."	
RL Appl. Environ. Microbiol. 67:4206-4214(2001).	
DR EMBL; AF385141; AAK70425.1; -	
SQ SEQUENCE 273 AA; 30864 MW; B4635B614DB9F3B3 CRC64;	
Query Match	100.0%; Score 34; DB 2; Length 273;
Best Local Similarity	100.0%; Pred. No. 8.7;
Matches 7; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LONLVIP 7	
DB 223 LONLVIP 229	
RESULT 2	
Q93DP7	PRELIMINARY; PRT; 273 AA.
AC Q93DP7;	
DT 01-DEC-2001 (TREMUREL. 19, Created)	
DT 01-DEC-2001 (TREMUREL. 19, Last sequence update)	
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)	

DE 2,5-DIKETO-D-GLUCONIC ACID REDUCTASE.
 OS uncultured bacterium.
 OC Bacteria; environmental samples.
 OX NCBI_Taxid=77133;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21417047; PubMed=11526025;
 RX Eschenfeldt W.H., Stols L., Rosenbaum H., Khambatta Z.S.,
 RA Quate-Randall E., Wu S., Kilgore D.C., Trent J.D., Donnelly M.I.;
 RT "Dna from uncultured organisms as a source of 2,5-diketo-D-gluconic
 RT acid reductases.";
 RL APPL. Environ. Microbiol. 67:4206-4214(2001).
 DR EMBL: AF385143; AAK70428.1; .
 SQ SEQUENCE 273 AA; 30879 MW; 8E73501027A93363 CRC64;

Query Match 100.0%; Score 34; DB 2; Length 273;
 Best Local Similarity 100.0%; Pred. No. 8.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
 |||||
 Db 223 LQNLVIP 229

RESULT 3
 ID Q9U943 PRELIMINARY; PRT; 3380 AA.
 AC Q9U943;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE APOLIPOPHORIN PROTEIN PRECURSOR.
 OS Locusta migratoria (Migratory locust).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
 OC Acridomorpha; Acridoidea; Acrididae; Locusta.
 OX NCBI_Taxid=7004;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FAT BODY;
 RA Bogerd J., Babin P.J., Koolman F.P., Van der Horst D.J.;
 RT "Cloning and expression of cDNA encoding the apolipoprotein precursor
 RT protein mRNA of Locusta migratoria.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ130944; CAB51918.1; .
 DR InterPro: IPR003662; sub_transporter.
 DR InterPro: IPR001747; Vitellogenin_N.
 DR InterPro: IPR001846; Vwd.
 DR Pfam: PF01347; Vitellogenin_N; 3.
 DR Pfam: PF00094; Vwd; 1.
 DR SMART: SM00216; VWD; 1.
 DR PROSITE: PS00217; SUGAR_TRANSPORT_2; UNKNOWN_N_1.
 KW Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 716 APOLIPOPHORIN II.
 FT CHAIN 721 3380 APOLIPOPHORIN I.
 SQ SEQUENCE 3380 AA; 371719 MW; DF3EE9B46EB8BDB5 CRC64;

Query Match 94.1%; Score 32; DB 5; Length 3380;
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
 |||||
 Db 2670 LQNLVIP 2676

RESULT 4
 ID Q9W616 PRELIMINARY; PRT; 358 AA.
 AC Q9W616;

DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE SIGNALING MOLECULE LEFTY1.
 GN LEFTY.
 OS Brachydanio rerio (zebrafish) (zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Danio.
 OX NCBI_Taxid=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99307087; PubMed=10375514;
 RA Bisgrove B.W., Essner J.J., Yost H.J.;
 RT "Regulation of midline development by antagonism of lefty and nodal
 RT signaling.";
 RL Development 126:3253-3262(1999).
 CC -I- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 DR EMBL: AF132444; AAD34388.1; .
 DR HSSP: P01137; IKLA.
 DR ZFIN: ZDB-GENE-990630-10; lft1.
 DR InterPro: IPR001839; TGF-beta.
 DR InterPro: IPR001111; TGFb_N.
 DR Pfam: PF000019; TGF-beta; 1.
 DR Pfam: PF00688; TGFb_propeptide; 1.
 DR ProDom: PD000357; TGF-beta; 1.
 DR SMART: SM00204; TGFb; 1.
 DR PROSITE: PS00250; TGF_BETA; 1.
 KW Glycoprotein.
 SQ SEQUENCE 358 AA; 41018 MW; F7A86FE828092A8 CRC64;

Query Match 91.2%; Score 31; DB 13; Length 358;
 Best Local Similarity 85.7%; Pred. No. 56;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
 |||||
 Db 47 LQNLVIP 53

RESULT 5
 ID Q9PW55 PRELIMINARY; PRT; 358 AA.
 AC Q9PW55;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ANTIYIN.
 GN LEFTY.
 OS Brachydanio rerio (zebrafish) (zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Danio.
 OX NCBI_Taxid=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99065494; PubMed=9847237;
 RA Thisse C., Thisse B.;
 RT "Antiyn, a novel and divergent member of the Tgfbeta superfamily,
 RT negatively regulates mesoderm induction.";
 RL Development 126:229-240(1999).
 CC -I- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 DR EMBL: AF172089; AAD47114.1; .
 DR HSSP: P01137; IKLA.
 DR ZFIN: ZDB-GENE-990630-10; lft1.
 DR InterPro: IPR001839; TGF-beta.
 DR InterPro: IPR001111; TGFb_N.
 DR Pfam: PF000019; TGF-beta; 1.
 DR Pfam: PF00688; TGFb_propeptide; 1.
 DR ProDom: PD000357; TGF-beta; 1.
 DR SMART: SM00204; TGFb; 1.
 DR PROSITE: PS00250; TGF_BETA; 1.

KW Glycoprotein.
SQ SEQUENCE 358 AA; 41041 MW; 74D3036F18D2458F CRC64;

Query Match 91.2%; Score 31; DB 13; Length 358;
Best Local Similarity 85.7%; Pred. No. 56;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVTP 7
|:|||||
DB 47 LQNLVTP 53

```

RESULT 6
P89442 PRELIMINARY; PRT; 1374 AA.
AC P89442;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MAJOR CAPSID PROTEIN.
GN UL19.
OS Herpes simplex virus (type 2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=87111457; PubMed=3027242;
RA McGeoch D.J., Moss H.W., McNab D., Frame M.C.;
RT "DNA sequence and genetic content of the HindIII 1 region in the short
RT unique component of the herpes simplex virus type 2 genome:
RT identification of the gene encoding glycoprotein G, and evolutionary
RT comparisons."
RT J. Gen. Virol. 68:19-38(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=90278430; PubMed=2161906;
RA Everett R., Fenwick M.;
RT "Comparative DNA sequence analysis of the host shutoff genes of
RT different strains of herpes simplex virus: type 2 strain HG52 encodes
RT a truncated UL41 product."
RT J. Gen. Virol. 71:1387-1390(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=92113549; PubMed=1662697;
RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT "Comparative sequence analysis of the long repeat regions and
RT adjoining parts of the long unique regions in the genomes of herpes
RT simplex viruses types 1 and 2."
RT J. Gen. Virol. 72:3057-3073(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=92356101; PubMed=1322965;
RA Barnett B.C., Dolan A., Telford E.A.R., Davison A.J., McGeoch D.J.;
RT "A novel herpes simplex virus gene (UL49A) encodes a putative membrane
RT protein with counterparts in other herpesviruses."
RT J. Gen. Virol. 73:2167-2171(1992).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RA Dolan A.;
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z86099; CAB06743.1; -
DR InterPro: IPR000912; Herpes_MCP.
DR Pfam: PF03122; Herpes_MCP. 1.
DR PRINTS: PR00235; HSVCAPSIDMCP.
SQ SEQUENCE 1374 AA; 149237 MW; EC8E85351E4C1939 CRC64;
```

Query Match 91.2%; Score 31; DB 12; Length 1374;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVTP 7
|:|||||
DB 839 LQNLVTP 845

```

RESULT 7
Q9WHF8 PRELIMINARY; PRT; 225 AA.
AC Q9WHF8;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE REPLICATION-ASSOCIATED PROTEIN (FRAGMENT).
GN REP.
OS Sinusioa tomato leaf curl geminivirus.
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=71186;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CR2;
RA Karakashian J.P., Nakha M.K., Maxwell D.P., Ramirez P.;
RT "Molecular characterization of tomato-infecting geminiviruses in Costa
RT Rica."
RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF131213; AAD34731.1; -
DR InterPro: IPR001191; Geminl_AL1.
DR Pfam: PF00799; Geminl_AL1; 1.
DR PRINTS: PR00227; GEMCOATALL.
DR ProDom: PD000736; Geminl_AL1; 1.
FT NON_TER 225
SQ SEQUENCE 225 AA; 25786 MW; 543C268BD052BDC8 CRC64;
```

Query Match 88.2%; Score 30; DB 12; Length 225;
Best Local Similarity 85.7%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 LQNLVTP 7
|:|||||
DB 32 LQNLVTP 38

```

RESULT 8
O39178 PRELIMINARY; PRT; 235 AA.
AC O39178;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AC1 (FRAGMENT).
OS Geminiviridae.
OC Viruses; ssDNA viruses.
OX NCBI_TaxID=10811;
RN [1]
RP SEQUENCE FROM N.A.
RA Guzman P., Arredondo C.R., Emmatty D., Portillo R.J., Gilbertson R.L.;
RT "Partial Characterization of Two Whitefly-Transmitted Geminiviruses
RT Infecting Tomatoes in Venezuela."
RT Plant Dis. 81:312-312(1997).
RL EMBL: AF026464; AAB82598.1; -
DR InterPro: IPR001191; Geminl_AL1.
DR Pfam: PF00799; Geminl_AL1; 1.
DR PRINTS: PR00227; GEMCOATALL.
DR ProDom: PD000736; Geminl_AL1; 1.
FT NON_TER 235
SQ SEQUENCE 235 AA; 26522 MW; 317E4D370F16D2DB CRC64;
```

```
Query Match      88.2%; Score 30; DB 12; Length 235;
Best Local Similarity 85.7%; Pred. No. 64;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 QNLVLP 7
      111111
Db      32 QNLVLP 38

RESULT 9
ID 021187 PRELIMINARY; PRT; 346 AA.
AC 021187;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 2 (FRAGMENT).
GN NADH2.
OS Lama nasus.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Galeoidea; Lamniformes; Alopiidae;
OC Lama.
OX NCBI_TaxID=7849;
RN [1]
RP SEQUENCE FROM N.A.
RA Naylor G.J.P., Martin A.P., Mattison E.G., Brown W.M.;
RT "Interrelationships of Lamniform Sharks: Testing Phylogenetic
   Hypotheses with Sequence Data.";
RL (In) Kocher T.D., Stepien C.A. (eds.);
   MOLECULAR SYSTEMATICS OF FISHES, pp.195-214, Academic Press,
   San Diego, CA, USA (1997).
CC -1- CATALYTIC ACTIVITY: NADH + URICINONE -> NAD(+) + URICINOL.
DE EMBL: U91427; AAB63136.1; -.
DR InterPro: IPR001750; Oxidored_q1.
PF pfam: PF00361; oxidored_q1; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
FT NON_TER
SQ SEQUENCE 346 AA; 38035 MW; F0A152FF19926F6 CRC64;

Query Match      88.2%; Score 30; DB 8; Length 346;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 QNLVLP 7
      111111
Db      273 QNLVLP 278

RESULT 10
ID 09W617 PRELIMINARY; PRT; 362 AA.
AC 09W617;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE SIGNALING MOLECULE LEFT2.
GN Brachydanio relio (zebrafish) (zebra danio).
OS Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99307087; PubMed=10375514;
RA Blagrove B.W., Essner J.J., Yost H.J.;
RT "Regulation of midline development by antagonism of lefty and nodal
   signaling.";
RL Development 126:3253-3262(1999).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
DE EMBL: AF132445; AAD34389.1; -.

+ 37
```

```
DR HSP: P08112; 2TGT.
DR ZFIN: ZDB-GENE-990630-11; lft2.
DR InterPro: IPR001839; TGF-beta.
DR InterPro: IPR001111; TGF_N.
DR Pfam: PF00019; TGF-beta; 1.
DR Pfam: PF00688; TGF-propeptide; 1.
DR ProDom: PD000357; TGF-beta; 1.
DR SMART: SM00204; TGFb; 1.
DR PROSITE: PS00250; TGF-BETA; 1.
KW Glycoprotein.
SQ SEQUENCE 362 AA; 41459 MW; 0C12B0CB3CEB6007 CRC64;

Query Match      88.2%; Score 30; DB 13; Length 362;
Best Local Similarity 71.4%; Pred. No. 96;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 QNLVLP 7
      111111
Db      46 QNLVLP 52

RESULT 11
ID 09PYV2 PRELIMINARY; PRT; 372 AA.
AC 09PYV2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE ORF91.
OS Xestia c-nigrum granulosis virus (Xncv) (Xestia c-nigrum
   granulovirus).
OS Virusess; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
OX NCBI_TaxID=51677;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99434230; PubMed=10502508;
RA Hayakawa T., Ko R., Okano K., Seong S.I., Goto C., Maeda S.;
RT "Sequence analysis of the Xestia c-nigrum granulovirus genome.";
RL Virology 262:277-297(1999).
DE EMBL: AF162221; AAF05205.1; -.
SQ SEQUENCE 372 AA; 43839 MW; ECE90B7BB79FA772 CRC64;

Query Match      88.2%; Score 30; DB 12; Length 372;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 QNLVLP 7
      111111
Db      338 QNLVLP 343

RESULT 12
ID 09CXB2 PRELIMINARY; PRT; 373 AA.
AC 09CXB2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE 8430417G17RK PROTEIN.
GN 8430417G17RK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawal J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
```


RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kato K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochant W.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schiraldi L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
 RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL NATURE 409:685-690(2001).
 DR EMBL, AK018423, BAB31203.1; -
 DR MGI:1923000, 8430417G17RLK.
 DR InterPro: IPR002657; SBF.
 DR Pfam: PF01758; SBF; 1.
 SO SEQUENCE 373 AA; 40681 MW; 0902D18506A8AC55 CRC64;

Query Match 88.2%; Score 30; DB 11; Length 373;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ONLVIP 7
 Db 156 ONLVIP 161

RESULT 13
 044116 PRELIMINARY; PRT; 634 AA.
 AC 044116;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VIMAR.
 GN VIMAR OR CG3572
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lo P.C.H., Frasch M.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF034421; AAB87984.1; -
 DR FLYBase: FBgn0022960; Vimar.
 DR InterPro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 5.
 DR SMART: SM00185; ARM; 3.
 SO SEQUENCE 634 AA; 70237 MW; 0968BD62DF1455FD CRC64;

Query Match 88.2%; Score 30; DB 5; Length 634;
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ONLVIP 7
 Db 385 LRLNVP 391

RESULT 14
 09V978 PRELIMINARY; PRT; 634 AA.
 ID 09V978;
 AC 09V978;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VIMAR PROTEIN.
 GN VIMAR OR CG3572.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Beeson K.Y., Bess A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Bertozzi D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaisai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson S.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003790; AAF57418.1; -
 DR FLYBase: FBgn0022960; Vimar.
 DR InterPro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 5.
 DR SMART: SM00185; ARM; 3.
 SO SEQUENCE 634 AA; 70253 MW; 337A23DB99A5486B CRC64;

Query Match 88.2%; Score 30; DB 5; Length 634;
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ONLVIP 7
 Db 385 LRLNVP 391

RESULT 15
 018668 PRELIMINARY; PRT; 689 AA.
 ID 018668;
 AC 018668;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE CA7DI2.2 PROTEIN.
GN CA7DI2.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gajadaty S.;
RU Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z69902; CAA93767.1; -;
SQ SEQUENCE 689 AA; 78202 MW; 8CF57ED089C0D49F CRC64;

Query Match 88.2%; Score 30; DB 5; Length 689;
Best Local Similarity 71.4%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVP 7
|:|:|:|
Db 392 LONLVP 398

Search completed: August 20, 2002, 11:31:56
Job time: 1470 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:07:17 ; Search time 187.18 Seconds
(without alignments)
2.967 Million cell updates/sec

Title: US-09-824-286-16

Perfect score: 34

Sequence: 1 HCLEH 5

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	34	100.0	5	19	AAW31653	Human cytokine rec
2	34	100.0	105	22	AAW31653	Human cytokine rec
3	34	100.0	174	19	AAW38627	Streptococcus pneu
4	34	100.0	192	18	AAW11279	Streptococcus pneu
5	34	100.0	229	21	AAW35440	Pinus radiata cell
6	34	100.0	230	15	AAW47151	IL-2 receptor gamm
7	34	100.0	230	16	AAW82934	Interleukin 4 comp
8	34	100.0	252	15	AAW47150	IL-2 receptor gamm
9	34	100.0	270	21	AAW54733	Arabidopsis thalia
10	34	100.0	297	21	AAW81769	Streptococcus pneu
11	34	100.0	347	15	AAW47149	IL-2 receptor gamm

12	34	100.0	363	22	AAW31682	Bos taurus interle
13	34	100.0	369	15	AAW47148	IL-2 receptor gamm
14	34	100.0	482	19	AAW31646	Human cytokine rec
15	34	100.0	691	21	AAW92202	Fusion polypeptide
16	34	100.0	694	21	AAW92201	Fusion polypeptide
17	34	100.0	694	21	AAW92203	Fusion polypeptide
18	34	100.0	2645	22	AAW20077	Human human diagno
19	32	94.1	105	22	AAW47409	Human colon cancer
20	32	94.1	364	22	AAW60565	Drosophila melanog
21	31	91.2	30	22	AAW79026	Human protein SEQ
22	31	91.2	56	22	AAW80010	Human protein SEQ
23	31	91.2	89	21	AAW33478	Arabidopsis thalia
24	31	91.2	141	21	AAW33477	Arabidopsis thalia
25	31	91.2	148	22	AAW53000	Propionibacterium
26	31	91.2	178	21	AAW18383	Arabidopsis thalia
27	31	91.2	199	11	AAW06498	GST-1 clone encode
28	31	91.2	270	21	AAW18382	Arabidopsis thalia
29	31	91.2	270	21	AAW49119	Arabidopsis thalia
30	31	91.2	317	21	AAW50357	Arabidopsis thalia
31	31	91.2	326	21	AAW50356	Arabidopsis thalia
32	31	91.2	328	21	AAW50355	Arabidopsis thalia
33	31	91.2	330	21	AAW49118	Arabidopsis thalia
34	31	91.2	338	21	AAW18381	Arabidopsis thalia
35	31	91.2	338	21	AAW49117	Arabidopsis thalia
36	31	91.2	425	14	AAW45151	Sequence of mouse
37	31	91.2	425	17	AAW06798	Murine p154. Mus
38	31	91.2	515	21	AAW42522	Arabidopsis thalia
39	31	91.2	748	22	AAW07878	Polypeptide sequen
40	31	91.2	1316	22	AAW39108	Human polypeptide
41	31	91.2	1401	22	AAW40894	Human polypeptide
42	31	91.2	1612	22	AAW71504	Drosophila melanog
43	30	88.2	75	22	AAW24193	Human EST encoded
44	30	88.2	84	22	AAW88953	Human immune/haema
45	30	88.2	141	22	AAW17027	Novel human diagno

ALIGNMENTS

RESULT 1	
AAW31653	
ID AAW31653 standard; Peptide: 5 AA.	
XX	
AC AAW31653:	
XX	
DT 21-MAY-1998 (first entry)	
XX	
DE Human cytokine receptor gc chain epitope.	
XX	
KW Cytokine receptor; gamma common chain; gc chain; human;	
KW blocking agent; monoclonal antibody; Cp.B8; immunological disease;	
KW myasthenia gravis; Rheumatoid arthritis; lupus; multiple sclerosis;	
KW insulin-dependent diabetes; inflammatory bowel disease;	
KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;	
KW graft versus host disease; psoriasis; immunosuppressive; therapy;	
KW epitope.	
XX	
OS Homo sapiens.	
XX	
PN WO9743416-A1.	
XX	
PD 20-NOV-1997.	
XX	
PF 09-MAY-1997; 97WO-US07870.	
XX	
PR 10-MAY-1996; 96US-0017466.	
XX	
PA (BIOJ) BIOGEN INC.	
XX	
PI Benjamin CD, Burkly LC, Hession C, Whitty A;	
XX	
DR WPI; 1998-008885/01.	
XX	

PT Blocking agents of the gamma common chain of cytokine receptors -
PT particularly monoclonal antibodies, used to induce T cell anergy for
PT treatment of immunological diseases
XX
XX
PS Claim 24; Page 85; 11pp; English.
XX
XX This peptide comprises an epitope of the human cytokine receptor
CC common gamma (gc) chain (see AAW31646) that is recognised by
CC gc blocking agents of the invention. 5 Such epitopes (see
CC AAW31650-54) have been identified. The invention provides
CC compositions and methods for inhibiting cytokine signalling using
CC gc chain blocking agents for the treatment of immunological
CC diseases such as myasthenia gravis, rheumatoid arthritis, lupus,
CC multiple sclerosis, insulin-dependent diabetes, inflammatory bowel
CC disease, sympathetic ophthalmia, uveitis, allergy, asthma,
CC parasitic infection, graft vs. host disease or psoriasis. A
CC preferred gc blocking agent is Mab CP.B8 or its Fab fragment (see
CC also AAW31647-48).
XX
SQ Sequence 5 AA;

Query Match 100.0%; Score 34; DB 19; Length 5;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
| | | | |
DB 1 hcleh 5

RESULT 2
AAG77373
ID AAG77373 standard; Protein; 105 AA.
XX
AC AAG77373;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SPQ ID NO:8139.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI: 2001-223537/24.

XX N-PSDB: AAW36780.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 9461-9462; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAW77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 105 AA;

Query Match 100.0%; Score 34; DB 22; Length 105;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
| | | | |
DB 9 hcleh 13

RESULT 3
AAW38627
ID AAW38627 standard; Protein; 174 AA.
XX
XX AAW38627;

XX 09-NOV-1998 (first entry)

XX Streptococcus pneumoniae protein of unknown function.

XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KW pathogenesis.

XX Streptococcus pneumoniae.

XX WO9743303-A1.

XX 20-NOV-1997.

XX 14-MAY-1997; 97WO-US07950.

XX 14-MAY-1996; 96US-0017670.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;

XX Stodola RK;

XX WPI: 1998-008793/01.

XX N-PSDB: AAT98673.

XX Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections
XX
PS Claim 12; Pages 387-388; 483pp; English.

XX This sequence represents a Streptococcus pneumoniae protein of
CC unknown function, and is encoded by a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 0100993 (NCIMB 40794). The streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be

CC used to treat diseases caused by *S. pneumoniae* proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the *S. pneumoniae* proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.

XX
SQ Sequence 174 AA;

Query Match 100.0%; Score 34; DB 19; Length 174;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
|||||
Db 91 hcleh 95

RESULT 4

ID AAY11279 standard; Protein; 192 AA.

XX AC AAY11279;

XX DT 20-MAY-1999 (first entry)

XX DE Streptococcus pneumoniae protein sequence ID NO:389.

XX KM Streptococcus pneumoniae strain 0100993; vaccine; immune response;
XX KW streptococcal infection; pneumococcal.

XX OS Streptococcus pneumoniae.

XX PN W09737026-A1.

XX PD 09-OCT-1997.

XX PF 01-APR-1997; 97WO-US05306.

XX PR 22-AUG-1996; 96US-0025788.

XX PR 02-APR-1996; 96US-0014690.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;

XX PT Stodola RK;

XX DR WPI; 1997-503111/46.

XX DR N-PSDB; AAX30862.

XX PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in
XX vaccines, drug screening, etc

XX PT Claim 6; Page 294-295; 354pp; English.

XX CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from
CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
CC encode the novel proteins given in AAY1114 to AAY11367. The proteins,
CC isolated from Streptococcus pneumoniae, can be used in vaccines against
CC streptococcal infections and in assays for identifying compounds that
CC inhibit or activate the activity of the proteins. The antagonists can
CC be used to treat an individual having need to inhibit a bacterial
CC protein. Vectors expressing the proteins can be used to induce a
XX protective immune response in mammals.

SQ Sequence 192 AA;

Query Match 100.0%; Score 34; DB 18; Length 192;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
|||||
Db 91 hcleh 95

RESULT 5

ID AAB25440 standard; Protein; 229 AA.

XX AC AAB25440;

XX DT 27-NOV-2000 (first entry)

XX DE Pinus radiata cell signalling involved protein smg ID NO:759.

XX KM Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
XX KW plant cell signalling; modulation; transgenic plant; pathogen; growth;
XX KW environmental change; development; cell proliferation; differentiation;
XX KW elongation; survival; disease resistance; nutrient metabolism.

XX OS Pinus radiata.

XX PN W0200042171-A1.

XX PD 20-JUL-2000.

XX PF 11-JAN-2000; 2000WO-US00724.

XX PR 12-JAN-1999; 99US-0228986.

XX PR 01-NOV-1999; 99US-0162866.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.

XX PI Strabala TJ, Nieuwenhuizen NJ;

XX DR WPI; 2000-476052/41.

XX PT Isolated polynucleotide encoding a polypeptide involved in cell
XX signaling used for generating transgenic plants with modified responses
XX to external signals -

XX PS Claim 3; Page 351-352; 527pp; English.

XX CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide
CC and protein sequences isolated from eucalyptus (*Eucalyptus grandis*) or
CC pine (*Pinus radiata* also known as Monterey pine). The protein sequences
CC are involved in cell signalling. The polynucleotide and protein
CC sequences can be used to modify the response of plant cells to external
CC signals e.g. environmental changes or pathogens during the growth and
CC development of a plant. They can be used to modify cell proliferation,
CC differentiation, elongation and survival, resistance to disease and
CC nutrient metabolism. Examples of modifications which can be produced are
CC altered fruit ripening and senescence of leaves and flowers e.g. to
CC delay senescence and prolong the life of cut flowers or enhance
CC senescence of reproductive organs to engineer sterile plants. Other
CC modifications can be used to delay senescence in selected cell types or
CC organs providing fruit and vegetables which have a longer shelf life
CC between harvest and consumption, or to decrease branching frequency in
CC forest tree species giving long stretches of valuable knot-free clear
XX wood which can be used in solid timber furniture and veneers.

SQ Sequence 229 AA;

Query Match 100.0%; Score 34; DB 21; Length 229;
Best Local Similarity 100.0%; Pred. No. 94;

PD 19-JAN-1994.
XX
PF 22-APR-1993; 93EP-0106561.
XX
PR 23-APR-1992; 92JP-0104947.
XX
PA (AJIN) AJINOMOTO KK.
PA (SUGA/) SUGAMURA K.
XX
PI Aseo H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
DR WPI; 1994-017546/03.
DR N-PSDB; AA054830.
XX
PT DNA and protein sequences of IL-2 gamma chain - useful as immune
PT regulatory agents for treatment of e.g. rheumatoid arthritis and
PT transplant rejection
XX
PS Disclosure; Page 21-22, 34-35; 50pp; English.
XX
CC The human IL-2 receptor gamma chain preform (AAR47148), including the
CC signal peptide, is encoded by the sequence given in AA054828. The
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
CC while a soluble form suitable for expression in prokaryotes (AAR47151)
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
CC terminal sequence of IL-2 receptor gamma chain, and are used to
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
CC are used to obtain the protein given in AAR47151.
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 34; DB 15; Length 252;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
Db 181 hcleh 185

RESULT 9
AA054733
ID AA054733 standard; Protein: 270 AA.
XX
AC AA054733;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 69867.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.

```

PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144332.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145386.
PR 02-AUG-1999; 99US-0145388.
PR 03-AUG-1999; 99US-0145389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158366.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.

```

```

PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 100.0%; Score 34; DB 21; Length 270;
 Best Local Similarity 100.0%; Pred. No. 1,1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLFH 5
 DB 187 hclfh 191

```

RESULT 10
AAV81769
ID AAV81769 standard; Protein; 297 AA.
XX
AC AAV81769;
XX
DT 02-JUN-2000 (first entry)
XX
DE Streptococcus pneumoniae protein sequence ID127.
XX
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
kidney disease; diabetes; immunosuppressive disorder; otitis media;
pneumococcal septicaemia; sinusitis; meningitis; therapy.
XX
OS Streptococcus pneumoniae.
XX
WO200006738-A2.
XX
PD 10-FEB-2000.
XX
PF 27-JUL-1999; 99WO-GB02452.
XX
PR 27-JUL-1998; 98GB-0016336.
PR 19-MAR-1999; 99US-0125329.
XX
PA (MICR-) MICROBIAL TECHNIQS LTD.
XX
PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
XX
DR WPI; 2000-195301/17.
DR N-PSDB; AA291865.
XX
PT Streptococcal proteins and polynucleotides useful for diagnosis,
treatment and prophylaxis of bacterial infections
XX
PS Claim 2; Page 63; 76pp; English.

```


XX This sequence represents a Streptococcus pneumoniae protein of the
 CC invention. The proteins (or their homologues, derivatives and/or
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
 CC compositions comprising the proteins are useful as vaccines and also in
 CC diagnostic assays. The sequences are useful for the detection or
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
 CC with them. Agents capable of antagonising, inhibiting or interfering with
 CC the function or expression of the protein or polypeptide are useful in
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae
 CC infection. As the sequences can be used to treat S. pneumoniae infection,
 CC they can be used to treat bacterial pneumonia, which has high rates in
 CC young children, the elderly, and in patients with predisposing conditions
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
 CC or with immunosuppressive disorders, especially AIDS. They can also be
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and
 CC meningitis.

XX Sequence 297 AA;
 SQ

Query Match 100.0%; Score 34; DB 21; Length 297;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
 |||||
 Db 214 hcleh 218

RESULT 11
 AAR47149
 ID AAR47149 standard; Protein; 347 AA.
 XX
 AC AAR47149;
 DT 13-JUN-1994 (first entry)
 XX
 DE IL-2 receptor gamma chain.
 XX
 KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
 KW rheumatoid arthritis; transplant rejection; primer;
 KW polymerase chain reaction; PCR; amplification.
 XX
 OS Homo sapiens.
 XX
 PN EP578932-A.
 PD 19-JAN-1994.
 XX
 PF 22-APR-1993; 93EP-0106561.
 XX
 PR 23-APR-1992; 92JP-0104947.
 XX
 PA (AJIN) AJINOMOTO K.
 PA (SUGA) SUGAMURA K.
 PI Aseo H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
 PI Suzuki M, Takeshita T;
 DR WPI; 1994-017546/03.
 DR N-PSDB; AAO54829.
 XX
 PT DNA and protein sequences of IL-2 gamma chain - useful as immune
 PT regulatory agents for treatment of e.g. rheumatoid arthritis and
 PT transplant rejection
 XX
 PS Claim 4; Page 41; 50pp; English.
 CC The human IL-2 receptor gamma chain preform (AAR47148), including the
 CC signal peptide, is encoded by the sequence given in AAO54828. The
 CC mature protein (AAR47149) is encoded by sequence AAO54829. A soluble
 CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AAO54830,

CC while a soluble form suitable for expression in prokaryotes (AAR47151)
 CC is encoded by AAO54831. Primers 1-6 (AAO54820-25) are based on the N-
 CC terminal sequence of IL-2 receptor gamma chain, and are used to
 CC isolate IL2 receptor gamma chain receptor cDNA. Primers AAO54826-27
 CC are used to obtain the protein given in AAR47151.
 XX
 SQ Sequence 347 AA;
 SQ

Query Match 100.0%; Score 34; DB 15; Length 347;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
 |||||
 Db 159 hcleh 163

RESULT 12
 AAB71682
 ID AAB71682 standard; protein; 363 AA.
 XX
 AC AAB71682;
 DT 10-MAY-2001 (first entry)
 XX
 DE Bos taurus Interleukin-2 receptor gamma.
 XX
 KW Cytokine receptor common gamma chain like; CRGCL; human;
 KW tumours; infections; inflammatory; immune disorder;
 KW neurodegenerative; cardiovascular; disorder.
 XX
 OS Bos taurus.
 XX
 PN WO200112672-A2.
 PD 22-FEB-2001.
 XX
 PF 17-AUG-2000; 2000WO-US22493.
 XX
 PR 18-AUG-1999; 99US-0376430.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA Ruben SA, Rosen CA, Moore PA;
 PI
 PI
 PI
 DR WPI; 2001-147547/15.
 DR
 XX
 PT New nucleic acid molecule encoding a human cytokine receptor common
 PT gamma chain like polypeptide, useful for treating, preventing and/or
 PT diagnosing e.g. tumors, inflammatory diseases and immunodeficiency
 PT conditions -
 XX
 PS Disclosure; Fig 2; 288pp; English.
 PS
 XX
 CC The present invention relates to a human cytokine receptor
 CC common gamma chain like protein (CRGCL). The invention is useful
 CC for treating, preventing and/or diagnosing conditions such
 CC as tumours, infections, inflammatory diseases,
 CC immune disorders, neurodegenerative disorder and cardiovascular
 CC disorder.
 CC
 SQ Sequence 363 AA;
 SQ

Query Match 100.0%; Score 34; DB 22; Length 363;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
 |||||
 Db 172 hcleh 176

RESULT 13
AAR47148
ID AAR47148 standard; Protein: 369 AA.
XX
AC AAR47148;
XX
DT 13-JUN-1994 (first entry)
XX
DE IL-2 receptor gamma chain.
XX
KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KW rheumatoid arthritis; transplant rejection; primer; PCR;
KW polymerase chain reaction; amplification; ss.
XX
OS Homo sapiens.
XX
FT Key Location/Qualifiers
FT Peptide 1..22
FT /label= Sig_peptide
XX
PN EP578932-A.
XX
PD 19-JAN-1994.
XX
PE 22-APR-1993; 93EP-0106561.
XX
PR 23-APR-1992; 92JP-0104947.
XX
PA (AJIN) AJINOMOTO KK.
PA (SUGA) SUGAMURA K.
XX
PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
DR N-PSDB: AAQ54828.
XX
DR WPI: 1994-017546/03.
XX
PT DNA and protein sequences of IL-2 gamma chain - useful as immune
PT regulatory agents for treatment of e.g. rheumatoid arthritis and
PT transplant rejection
XX
PS Disclosure: Page 16-17, 29-30; 50pp; English.
XX
CC The human IL-2 receptor gamma chain preform (AAR47148), including the
CC signal peptide, is encoded by the sequence given in AAQ54828. The
CC mature protein (AAR47149) is encoded by sequence AAQ54829. A soluble
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AAQ54830.
CC While a soluble form suitable for expression in prokaryotes (AAR47151)
CC is encoded by AAQ54831. Primers 1-6 (AAQ54820-25) are based on the N-
CC terminal sequence of IL-2 receptor gamma chain, and are used to
CC isolate IL2 receptor gamma chain, receptor cDNA. Primers AAQ54826-27
CC are used to obtain the protein given in AAR47151.
XX
SQ Sequence 369 AA:

Query Match 100.0%; Score 34; DB 15; Length 369;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
DB 181 hcleh 185

RESULT 14
AAW31646
ID AAW31646 standard; Protein: 482 AA.
XX
AC AAW31646;
XX
DT 21-MAY-1998 (first entry)

XX
DE Human cytokine receptor gc chain-ig fusion protein.
XX
KW Blocking receptor; gamma common chain; gc chain; human;
KW myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
KW insulin-dependent diabetes; inflammatory bowel disease;
KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;
KW graft versus host disease; psoriasis; immunosuppressive; therapy.
XX
OS Chimeric - Homo sapiens.
XX
FT Key Location/Qualifiers
FT Protein 1..254
FT /note="gc chain N-terminal region"
FT Protein 255..482
FT /note="IgG1 constant region"
FT Region 255..264
FT /note="IgG1 hinge region"
FT Domain 264..482
FT /note="IgG1 CH2 and CH3 constant domains1"
XX
PN WO9743416-A1.
XX
PD 20-NOV-1997.
XX
PE 09-MAY-1997; 97WO-US07870.
XX
PR 10-MAY-1996; 96US-0017466.
XX
PA (BIOJ) BIOGEN INC.
XX
PI Benjamin CD, Burkly LC, Hession C, Whitty A;
PI WPI: 1998-008885/01.
XX
DR N-PSDB: AAT97439.
XX
PT Blocking agents of the gamma common chain of cytokine receptors -
PT particularly monoclonal antibodies, used to induce T cell energy for
PT treatment of immunological diseases
XX
PS Example 1; Page 79-80; 111pp; English.
XX
CC This polypeptide comprises a fusion between the N-terminal 254
CC amino acids of the human mature cytokine receptor gamma common (gc)
CC chain and the hinge region and CH2 and CH3 constant domains of
CC human IgG1. The fusion was expressed from clone PLB001 (see
CC AAT97439) in COS-7 cells, and used to generate murine anti-human gc
CC specific monoclonal antibodies (MAbs), including CP.B8 produced by
CC hybridoma ATCC HB 12107. The invention provides compositions and
CC methods for inhibiting cytokine signalling using gc chain blocking
CC agents for the treatment of immunological diseases such as
CC myasthenia gravis, rheumatoid arthritis, lupus, multiple sclerosis,
CC insulin-dependent diabetes, inflammatory bowel disease, sympathetic
CC ophthalmia, uveitis, allergy, asthma, parasitic infection, graft
CC vs. host disease or psoriasis. A preferred gc blocking agent is
CC Mab CP.B8 or its Fab fragment (see also AAW31647-48).
XX
SQ Sequence 482 AA:

Query Match 100.0%; Score 34; DB 19; Length 482;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
DB 181 hcleh 185

RESULT 15
AAV92202
ID AAV92202 standard; Protein: 691 AA.

Search completed: August 20, 2002, 11:07:18
 Job time: 6103 sec

```

XX AAY92202;
AC
XX
XX 01-AUG-2000 (first entry)
DT
XX
XX Fusion polypeptide 603, IL-4 trap.
DE
XX
XX IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;
KM cytostatic; immunomodulator; osteopathic.
XX
OS Synthetic.
OS Homo sapiens.
PN WO200018932-A2.
XX
XX 06-APR-2000.
PD
XX
XX 22-SEP-1999; 99WO-US22045.
PF
XX
XX 25-SEP-1998; 98US-0101858.
PR
XX 19-MAY-1999; 99US-0313942.
XX
XX (REG-) REGENERON PHARM INC.
PA
XX
XX Stahl N, Yancopoulos GD;
PI
XX
XX WPI: 2000-293165/25.
DR
XX N-PSDB; AAA09044.
XX
XX Isolated nucleic acid molecule for treating cytokine-related diseases
PT or disorders encodes a fusion polypeptide capable of binding a cytokine
PT to form a nonfunctional complex
XX
XX Example 6; Fig 22A-D; 152pp; English.
PS
XX
XX This sequence shows fusion polypeptide 603, which is capable of
CC binding cytokine IL-4 to form a non-functional complex.
CC
CC The invention concerns production of antagonists to any cytokine that
CC utilizes an alpha specificity determining component, which when combined
CC with the cytokine, binds to a first beta signal transducing component to
CC form a non-functional intermediate which then binds to a second beta
CC signal transducing component causing beta-receptor dimerization, the
CC soluble alpha specificity determining component of the receptor
CC (SR-alpha) and the extracellular domain of the first beta signal
CC transducing component of the cytokine receptor (beta-1) are combined to
CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the
CC cytokine by binding the cytokine to form a non-functional complex. The
CC receptor components are shared by cytokines such as the CNTF (cllary
CC neurotrophic factor) family of cytokines. The invention provides the
CC basis for the development of IL-6 antagonists, as they show that if, in
CC the presence of a ligand, a non-functional intermediate complex,
CC consisting of the ligand, its alpha receptor and its beta-1 receptor
CC component, can be formed, it will effectively block the action of the
CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers
CC of the extracellular domains of the alpha specificity determining
CC components of their receptors and the extracellular domain of gp130.
CC The resultant heterodimers, function as high-affinity traps, rendering
CC the cytokine inaccessible to form a signal transducing complex with the
CC native membrane-bound forms of their receptor. The nucleic acids and
CC polypeptides are useful for treating cytokine-related diseases or
CC disorders such as osteoporosis and primary and secondary effects of
CC cancer including multiple myeloma or cachexia.
XX
XX
SO Sequence 691 AA:

```

Query Match 100.0%; Score 34; DB 21; Length 691;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
 |||||
 Db 181 hcleh 185

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:10:37 ; Search time 100.84 Seconds

(without alignments)
4.764 Million cell updates/sec

Title: US-09-824-286-16

Perfect score: 34

Sequence: 1 HCLEH 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	129	2	AB1993
2	34	100.0	287	2	A95129
3	34	100.0	297	2	F97999
4	34	100.0	351	1	D84430
5	34	100.0	364	1	D86281
6	34	100.0	369	2	A42565
7	34	100.0	373	2	A55718
8	34	100.0	1044	2	T00342
9	34	100.0	2117	2	T36180
10	32	94.1	85	2	F82600
11	32	94.1	1829	2	TJ34239
12	31	91.2	30	2	JC1360
13	31	91.2	161	2	F72647
14	31	91.2	191	2	T23594
15	31	91.2	304	1	C84430
16	31	91.2	333	1	H86381
17	31	91.2	417	2	I50407
18	31	91.2	425	2	A46251
19	31	91.2	477	1	C2C2AM
20	31	91.2	515	2	B96825
21	31	91.2	632	2	D84921
22	31	91.2	694	2	G97400
23	31	91.2	746	2	AG2618
24	31	91.2	1314	2	G02870
25	31	91.2	2078	2	T25400
26	30	88.2	186	2	A11986
27	30	88.2	249	2	T46184
28	30	88.2	335	2	E87625
29	30	88.2	436	2	JC1497

ALIGNMENTS

30	29	85.3	75	2	D27393	11K inner spore co
31	29	85.3	119	2	AD3319	hypothetical prote
32	29	85.3	140	2	T16574	hypothetical prote
33	29	85.3	155	2	B95086	hypothetical prote
34	29	85.3	155	2	F97953	cytidine deaminase
35	29	85.3	165	2	F82136	conserved hypothet
36	29	85.3	186	2	AE1622	B. subtilis COMEB
37	29	85.3	187	2	AC1260	hypothetical prote
38	29	85.3	187	2	S75522	hypothetical prote
39	29	85.3	242	2	A96538	hypothetical prote
40	29	85.3	278	2	D70036	exopolysaccharide
41	29	85.3	299	2	S31771	HRPII protein - ma
42	29	85.3	301	2	S31782	HRPII protein - ma
43	29	85.3	301	2	S31773	HRPII protein - ma
44	29	85.3	308	2	T29317	hypothetical prote
45	29	85.3	309	2	A25942	histidine/alanine-

RESULT 1

AB1993 virulence associated protein C [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp. A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002

C:Accession: AB1993 R:Kanezo, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Irig

Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat

DNA Res. 8, 205-213, 2001

A:title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:reference number: AB1807, MID:21595285, PMID:11759840

A:Accession: AB1993

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-129 <KUR>

A:Cross-references: GB:BA000019; PIDN:BAH77861.1; PID:g17135316; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: vapC

C:Superfamily: virulence-associated protein vapC

Query Match 100.0%; Score 34; DB 2; Length 129;

Best Local Similarity 100.0%; Pred. No. 13;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5

Db 103 HCLEH 107

RESULT 2

A95129 transcription regulator MutR, probable [imported] - Streptococcus pneumoniae (strain

C:Species: Streptococcus pneumoniae C:date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: A95129 R:Teitelim, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; I

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple

nsou, T.; Hickey, E.K.; Holt, I.E. Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris

A:title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MID:21357209; PMID:11463916

A:Accession: A95129

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-287 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75226.1; PID:g14972591; GSPDB:GN00164; TIGR:

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP115

Query Match 100.0%; Score 34; DB 2; Length 287;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
|||||
DB 204 HCLEH 208

RESULT 3

F97999
transcription regulator rgpd [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
R:Accession: F97999
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Ee, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Meyer, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: F97999
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99826.1; PID:G15458639; GSPDB:GN00174
C:Genetics:
A:Gene: rgpd

Query Match 100.0%; Score 34; DB 2; Length 297;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
|||||
DB 214 HCLEH 218

RESULT 4

D84430
probable acid phosphatase (EC 3.1.3.2) At2g01890 precursor [similarity] - Arabidopsis thaliana
N:Alternate names: purple acid phosphatase
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Nov-2001
C:Accession: D84430
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; Euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: D84430
A:Molecule type: DNA
A:Residues: 1-351 <STO>
A:Cross-references: GB:AE002093; NID:94522012; PIDN:AAD21785.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g01890
A:Map position: 2
C:Superfamily: tartrate-resistant acid phosphatase; phosphoesterase core homology
C:Keywords: iron; metalloprotein; phosphoric monoester hydrolase
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-351/Product: acid phosphatase #status predicted <MAT>
F:46-125/Domain: phosphoesterase core homology <PEC>
F:52,65,88,270/Binding site: iron (Asp, Asp, Tyr, His) #status predicted
F:55,123,233,268/Binding site: iron (Asp, Asn, His, His) #status predicted
F:124,242/Active site: His #status predicted

Query Match 100.0%; Score 34; DB 1; Length 351;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
|||||
DB 270 HCLEH 274

RESULT 5

D86281
probable acid phosphatase (EC 3.1.3.2) F10B6.10 precursor [similarity] - Arabidopsis thaliana
N:Alternate names: purple acid phosphatase
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2001
R:Accession: D86281
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Lin, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maltl, R.; Marzla Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: D86281
A:Molecule type: DNA
A:Residues: 1-364 <STO>
A:Cross-references: GB:AE005172; NID:98778212; PIDN:AAK79221.1; GSPDB:GN00141
C:Genetics:
A:Gene: F10B6.10
A:Map position: 1
C:Superfamily: tartrate-resistant acid phosphatase; phosphoesterase core homology
C:Keywords: iron; metalloprotein; phosphoric monoester hydrolase
F:1-12/Domain: signal sequence #status predicted <SIG>
F:13-364/Product: acid phosphatase #status predicted <MAT>
F:75-152/Domain: phosphoesterase core homology <PEC>
F:81,114,117,281/Binding site: iron (Asp, Asp, Tyr, His) #status predicted
F:114,150,244,279/Binding site: iron (Asp, Asn, His, His) #status predicted
F:151,253/Active site: His #status predicted

Query Match 100.0%; Score 34; DB 1; Length 364;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
|||||
DB 281 HCLEH 285

RESULT 6

A42565
interleukin-2 receptor gamma chain - human
C:Species: Homo sapiens (man)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
R:Accession: A42565; A46591; I54332
R:Takeshita, T.; Asao, H.; Ohtani, K.; Ishii, N.; Kumaki, S.; Tanaka, N.; Munakata, H. Science 257, 379-382, 1992
A:Title: Cloning of the gamma chain of the human IL-2 receptor.
A:Reference number: A42565; MUID:92335883
A:Accession: A42565
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid; protein
A:Residues: 1-369 <TRK>
A:Cross-references: GB:D11086; NID:9303611; PIDN:BA01857.1; PID:9219890
A:Experimental source: MOLT beta lymphoid cells
A:Note: sequence extracted from NCBI backbone (NCBI:109167)
R:Noguchi, M.; Adelstein, S.; Cao, X.; Leonard, W.J.
J. Biol. Chem. 268, 13601-13608, 1993
A:Title: Characterization of the human Interleukin-2 receptor gamma chain gene.

A:Reference number: A46591; MUID:93293887
A:Accession: A46591
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369 <RES>
A:Cross-references: GB:U12183; NID:9307056; PIDN:AAA59145.1; PID:9307058
R:Puck, J.M.; Deschenes, S.M.; Porter, J.C.; Dutra, A.S.; Brown, C.J.; Willard, H.F.; He Hum. Mol. Genet. 2, 1099-1104, 1993
A:Title: The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-linked
A:Reference number: I54332; MUID:94004847
A:Accession: I54332
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369 <RE2>
A:Cross-references: GB:U19546; NID:9349631; PIDN:AAC37524.1; PID:9349632
C:Genetics:
A:Gene: GDB:IL2RG; SCIDX1; IMD4
A:Cross-references: GDB:134807; OMIM:308380
A:Map position: Xq13.1-Xq13.1
A:Introns: 39/1; 90/2; 152/1; 198/3; 253/1; 285/2; 308/3
A:Note: defects are associated with an X-linked form of severe combined immunodeficiency
C:Superfamily: interleukin-2 receptor gamma chain
C:Keywords: cytokine receptor; duplication; immunodeficiency; severe combined immunodeficiency

Query Match 100.0%; Score 34; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
|||||
Db 181 HCLEH 185

RESULT 7
A55718
Interleukin-2 receptor gamma chain precursor - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 23-Jul-1999
C:Accession: A55718
R:Henthorn, P.S.; Somberg, R.L.; Finlan, V.M.; Puck, J.M.; Patterson, D.F.; Felsburg, F
Genomics 23, 69-74, 1994
A:Title: IL-2Rgamma gene microdeletion demonstrates that canine X-linked severe combined
A:Reference number: A55718; MUID:95130114
A:Accession: A55718
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-373 <HEN>
A:Cross-references: GB:U04361; NID:9517411; PIDN:AAC48403.1; PID:9517412
C:Superfamily: interleukin-2 receptor gamma chain
C:Keywords: cytokine receptor; duplication

Query Match 100.0%; Score 34; DB 2; Length 373;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
|||||
Db 181 HCLEH 185

RESULT 8
T00342
hypothetical protein KIA0580 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00342
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545
A:Accession: T00342

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1044 <NAG>
A:Cross-references: EMBL:AB011152; NID:93043683; PIDN:BA25506.1; PID:93043684
A:Experimental source: brain; clone HD0601
C:Genetics:
A:Note: KIA0580

Query Match 100.0%; Score 34; DB 2; Length 1044;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
|||||
Db 954 HCLEH 958

RESULT 9
T36180
CDA peptide synthetase III - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000
C:Accession: T36180
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream
submitted to the EMBL data library, March 1999
A:Reference number: Z21600
A:Accession: T36180
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2117 <SAU>
A:Cross-references: EMBL:AL035707; PIDN:CAB38876.1; GSPDB:GN00070; SCOREDB:SCE63.01
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: caaP33; SCOREDB:SCE63.01
C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:208-662/Domain: acetate-CoA ligase homology <ACLI>
F:678-746/Domain: acyl carrier protein homology <ACPI>
F:1182-1725/Domain: acetate-CoA ligase homology <ACI2>
F:1140-1806/Domain: acyl carrier protein homology <ACR2>
F:1710-1772/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 100.0%; Score 34; DB 2; Length 2117;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
|||||
Db 466 HCLEH 470

RESULT 10
F82600
hypothetical protein XP2097 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82600
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Se
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: F82600
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-85 <SIM>
A:Cross-references: GB:AE004025; GB:AE003849; NID:99107217; PIDN:AAF84896.1; GSPDB:G
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvaranga, R
Bioness, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000
A:Authors: Perrella, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Klieger, J.E.; Kuramae, E.E.; Laiz
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B
A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
Rodrigues, V.; Rosa, A.J. de M.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.
A:Authors: da Silva, A.C.R.; da Silva, F.R.; de Sa, R.G.; Santelli, R.V.; Sawasak
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
A:Genetics:
A:Gene: XF2097

Query Match 94.1%; Score 32; DB 2; Length 85;
Best Local Similarity 80.0%; Pred. No. 21;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HCLEH 5
DB 15 HCLDH 19

RESULT 11
T34239
hypothetical protein F26F12.7 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34239
R:Wilson, R.; Bentley, D.; Gatling, S.
Submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F26F12.
A:Reference number: Z21493
A:Accession: T34239
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1829 <MIL>
A:Cross-references: EMBL:U55373; PIDN:AAC25894.1; GSPDB:GN00023; CESP:F26F12.7
C:Genetics:
A:Gene: CESP:F26F12.7
A:Map position: 5
A:Introns: 110/3; 441/3; 801/2; 1244/3; 1693/2; 1784/1

Query Match 94.1%; Score 32; DB 2; Length 1829;
Best Local Similarity 80.0%; Pred. No. 3,1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
DB 299 HCLDH 303

RESULT 12
JC1360
hypothetical 3k protein (HV3 5' region) - human
C:Species: *Homo sapiens* (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 28-May-1999
C:Accession: JC1360
R:Shimoto, T.; Tsuchimoto, H.; McGregor, C.G.A.; Mutch, H.; Shimizu, T.; Kurachi, Y.
Biochem. Biophys. Res. Commun. 189, 617-624, 1992
A:Title: Molecular cloning and characterization of the platelet-activating factor receptor
A:Reference number: JC1359; MUID:93112021
A:Accession: JC1360
A:Molecule type: mRNA
A:Residues: 1-30 <STUG>
A:Cross-references: GB:S52624; NID:g262469; PIDN:AAB24696.1; PID:g1680456
A:Experimental source: heart

Query Match 91.2%; Score 31; DB 2; Length 30;

Best Local Similarity 80.0%; Pred. No. 13;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HCLEH 5
DB 14 HCLDH 18

RESULT 13
F72647
hypothetical protein APE0612 - *Aeropyrum pernix* (strain K1)
C:Species: *Aeropyrum pernix*
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: F72647
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aero*
A:Reference number: A72450; MUID:99310339
A:Accession: F72647
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-161 <KAW>
A:Cross-references: DBJ:AP000060; NID:g5104188; PIDN:BAAT9582.1; PID:g5104267
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0612
C:Superfamily: *Aeropyrum pernix* hypothetical protein APE0612

Query Match 91.2%; Score 31; DB 2; Length 161;
Best Local Similarity 80.0%; Pred. No. 56;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HCLEH 5
DB 144 HCLDH 148

RESULT 14
T23594
hypothetical protein K10H10.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000
C:Accession: T23594
R:Percy, C.
Submitted to the EMBL Data Library, December 1996
A:Reference number: Z19766
A:Accession: T23594
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <MIL>
A:Cross-references: EMBL:Z83236; PIDN:CAB05781.1; GSPDB:GN00020; CESP:K10H10.5
A:Experimental source: clone K10H10
C:Genetics:
A:Gene: CESP:K10H10.5
A:Map position: 2
A:Introns: 22/2; 89/1
C:Superfamily: *Caenorhabditis elegans* hypothetical protein K10H10.4

Query Match 91.2%; Score 31; DB 2; Length 191;
Best Local Similarity 80.0%; Pred. No. 65;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
DB 41 HCLDH 45

RESULT 15
C84430
probable acid phosphatase (EC 3.1.3.2) At2g01880 precursor [similarity] - *Arabidopsis*

N:Alternate names: purple acid phosphatase
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Nov-2001
C:Accession: C84430
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: C84430
A:Molecule type: DNA
A:Residues: 1-304 <STO>
A:Cross-references: GB:AE002093; NID:g4522007; PIDN:AAD21780.1; GSPDB:GN00139
C:Genetics:
A:Gene: AT2G01880
A:Map position: 2
C:Superfamily: tartarate-resistant acid phosphatase; phosphoesterase core homology
C:Keywords: iron; metalloprotein; phosphoric monoester hydrolase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-304/Product: acid phosphatase #status predicted <MAT>
F:42-121/Domain: phosphoesterase core homology <PECC>
F:48,81,84,226/Binding site: iron (Asp, Asp, Tyr, His) #status predicted
F:81,119,189,224/Binding site: iron (Asp, Asn, His, His) #status predicted
F:120,198/Active site: His #status predicted

Query Match 91.2% Score 31; DB 1; Length 304;
Best Local Similarity 80.0% Pred. NO. 99;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 HCLEH 5
| | | | |
Db 226 HCLQH 230

Search completed: August 20, 2002, 11:10:39
Job time: 5669 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:33:02 : Search time 55.29 Seconds

(without alignments)
3.501 Million cell updates/sec

Title: US-09-824-286-16

Perfect score: 34

Sequence: 1 HCLEH 5

Scoring table: BLOSUM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	369	1	CYRG_HUMAN
2	34	100.0	373	1	CYRG_CANFA
3	34	100.0	379	1	CYRG_BOVIN
4	31	91.2	220	1	GT29_FASHE
5	31	91.2	416	1	PGK_CHICK
6	31	91.2	425	1	ADFP_MOUSE
7	31	91.2	477	1	GUNA_CLOTT
8	31	91.2	1314	1	Y197_HUMAN
9	31	91.2	1402	1	AFAM_MOUSE
10	29	85.3	75	1	COTD_BACSU
11	29	85.3	312	1	CDAL_YEAST
12	29	85.3	314	1	MHPB_ECOLI
13	29	85.3	326	1	INH5_ECOLI
14	29	85.3	332	1	HRP1_PLAFA
15	29	85.3	338	1	INSR_ECOLI
16	29	85.3	361	1	COL9_ARATH
17	29	85.3	450	1	ADFP_BOVIN
18	29	85.3	714	1	GREM_CHLAM
19	29	85.3	716	1	BGAL_THETU
20	29	85.3	742	1	ZFA_MOUSE
21	29	85.3	791	1	G6PE_HUMAN
22	29	85.3	805	1	DE19_CAEEL
23	29	85.3	870	1	POL_ISR
24	29	85.3	1313	1	MIRP_SCHPO
25	29	85.3	32	1	Y180_TREPA
26	28	82.4	61	1	TBA4_MAIZE
27	28	82.4	158	1	NUDB_HAEIN
28	28	82.4	189	1	REX3_MOUSE
29	28	82.4	242	1	PEPE_XENLA
30	28	82.4	323	1	RFC4_YEAST
31	28	82.4	360	1	CB2R_HUMAN
32	28	82.4	418	1	TBA_AJECA
33	28	82.4	418	1	TBA_AJECA

34	28	82.4	444	1	TBA_ONCKE	P30436 oncorhynch
35	28	82.4	447	1	TBA2_ELEIN	O22348 eleusine in
36	28	82.4	447	1	TBA_AVECA	P34690 caenorhabdi
37	28	82.4	448	1	TBA2_CABEL	P05215 homo sapien
38	28	82.4	448	1	TBA4_HUMAN	P09644 gallus gall
39	28	82.4	449	1	TBA5_CHICK	P24633 emericella
40	28	82.4	449	1	TBA1_EMENT	P53372 pneumocysti
41	28	82.4	449	1	TBA1_PNECA	P06604 drosophila
42	28	82.4	449	1	TBA2_DROME	P04689 schizosach
43	28	82.4	449	1	TBA2_SCHPO	P05218 mus musculu
44	28	82.4	449	1	TBA6_MOUSE	P11480 physarum po
45	28	82.4	449	1	TBAE_PHYPO	

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	369 AA.
CYRG_HUMAN				
ID	CYRG_HUMAN			
AC	P31785:			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cytokine receptor common gamma chain precursor (Gamma-C) (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (p64) (CD132 antigen).			
GN	IL2RG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RA	MEDLINE=92335883; PubMed=1631559;			
RA	Takeshita T., Asao H., Ohtani K., Ishii N., Kumaki S., Tanaka N.,			
RT	Munakata H., Nakamura M., Sugamura K.;			
RL	"Cloning of the gamma chain of the human IL-2 receptor.";			
RL	Science 257:379-382(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RC	MEDLINE=93293887; PubMed=8514792;			
RA	Noguchi M., Adelstein S., Cao X., Leonard W.J.;			
RT	"Characterization of the human Interleukin-2 receptor gamma chain			
RT	gene.";			
RL	J. Biol. Chem. 268:13601-13608(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND VARIANTS ASP-114 AND ASN-153.			
RA	MEDLINE=94004847; PubMed=8401490;			
RA	Puck J.M., Deschenes S.M., Porter J.C., Dutra A.S., Brown C.J.,			
RA	Willard H., Henchorn P.S.;			
RT	"The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated			
RT	in X-linked severe combined immunodeficiency, SCIDX1.";			
RL	Hum. Mol. Genet. 2:1099-1104(1993).			
RN	[4]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RA	MEDLINE=94090315; PubMed=8266076;			
RA	Kondo M., Takeshita T., Ishii N., Nakamura M., Watanabe S.,			
RT	Arai K.-I., Sugamura K.;			
RT	"Sharing of the interleukin-2 (IL-2) receptor gamma chain between			
RL	receptors for IL-2 and IL-4.";			
RL	Science 262:1874-1877(1993).			
RN	[5]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RA	MEDLINE=94090317; PubMed=8266078;			
RA	Russell S.M., Kegan A.D., Harada N., Nakamura Y., Noguchi M.,			
RA	Leonard W.J.,			
RT	"Interleukin-2 receptor gamma chain: a functional component of the			
RT	interleukin-4 receptor.";			
RL	Science 262:1880-1883(1993).			
RN	[6]			
RP	IDENTIFICATION AS A IL-7R SUBUNIT.			

RX MEDLINE-94090316; PubMed-8266077;
 RA Noguichi M., Nakamura Y., Russell S.M., Ziegler S.F., Tsang M., Cao X.,
 RA Leonard W.J.;
 RT "Interleukin-2 receptor gamma chain: a functional component of the
 RT Interleukin-7 receptor.";
 RL Science 262:1877-1880(1993).
 RN [17]
 RP 3D-STRUCTURE MODELING OF 57-248.
 RX MEDLINE-95111955; PubMed-7529123;
 RA Bamorough P., Hedgecock C.J., Richards M.G.;
 RT "The interleukin-2 and interleukin-4 receptors studied by molecular
 RT modelling";
 RL Structure 2:839-851(1994).
 RN [18]
 RP VARIANTS XSCID PHE-115; CYS-240 AND ILE-241.
 RX MEDLINE-94130970; PubMed-8299698;
 RA Disanto J.P., Dautry-Varsat A., Certain S., Fischer A.,
 RA de Saint Basile G.;
 RT "Interleukin-2 (IL-2) receptor gamma chain mutations in X-linked
 RT severe combined immunodeficiency disease result in the loss of
 RT high-affinity IL-2 receptor binding.";
 RL Eur. J. Immunol. 24:475-479(1994).
 RN [19]
 RP VARIANT XSCID LYS-68.
 RX MEDLINE-94375038; PubMed-8088810;
 RA Markiewicz S., Subtil A., Dautry-Varsat A., Fischer A.,
 RA de Saint Basile G.;
 RT "Detection of three nonsense mutations and one missense mutation in
 RT the Interleukin-2 receptor gamma chain gene in SCIDX1 that
 RT differentially affect the mRNA processing.";
 RL Genomics 21:291-293(1994).
 RN [10]
 RP VARIANT XSCID HIS-162.
 RX MEDLINE-94300093; PubMed-8027558;
 RA Ishii N., Aaso H., Kimura Y., Takeshita T., Nakamura M., Tsuchiya S.,
 RA Kono T., Maeda M., Uchiyama T., Sugamura K.;
 RT "Impairment of ligand binding and growth signaling of mutant IL-2
 RT receptor gamma-chains in patients with X-linked severe combined
 RT immunodeficiency.";
 RL J. Immunol. 153:1310-1317(1994).
 RN [11]
 RP VARIANT XSCID ASN-39.
 RX MEDLINE-95023932; PubMed-7937790;
 RA Disanto J.P., Rieux-Laucat F., Dautry-Varsat A., Fischer A.,
 RA de Saint Basile G.;
 RT "Defective human Interleukin 2 receptor gamma chain in an atypical X
 RT chromosome-linked severe combined immunodeficiency with peripheral T
 RT cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9466-9470(1994).
 RN [12]
 RP VARIANTS XSCID CYS-226 AND HIS-226.
 RX MEDLINE-95397841; PubMed-7668284;
 RA Pepper A.E., Buckley R.H., Small T.N., Puck J.M.;
 RT "Two mutational hotspots in the interleukin-2 receptor gamma chain
 RT gene causing human X-linked severe combined immunodeficiency.";
 RL Am. J. Hum. Genet. 57:564-571(1995).
 RN [13]
 RP VARIANT XSCID SER-183.
 RX MEDLINE-96013903; PubMed-7557965;
 RA Clark P.A., Lester T., Genet S., Jones A.M., Hendrike R.,
 RA Levisky R.L., Kinon C.;
 RT "Screening for mutations causing X-linked severe combined
 RT immunodeficiency in the IL-2R gamma chain gene by single-strand
 RT conformation polymorphism analysis.";
 RL Hum. Genet. 96:427-432(1995).
 RN [14]
 RP VARIANT XSCID GLN-237 G-H-N INS.
 RX MEDLINE-95164726; PubMed-7860773;
 RA Puck J.M., Pepper A.E., Bedard P.-M., Laframboise R.;
 RT "Female germ line mosaicism as the origin of a unique IL-2 receptor
 RT gamma-chain mutation causing X-linked severe combined
 RT immunodeficiency.";
 RL J. Clin. Invest. 95:895-899(1995).

RN [15]
 RP VARIANT XSCID GLN-293.
 RX MEDLINE-95190013; PubMed-7883965;
 RA Schmalstieg F.C., Leonard W.J., Noguichi M., Berg M., Rudloff H.E.,
 RA Denney R.M., Dave S.K., Brooks E.G., Goldman A.S.;
 RT "Missense mutation in exon 7 of the common gamma chain causes a
 RT moderate form of X-linked combined immunodeficiency.";
 RL J. Clin. Invest. 95:1169-1173(1995).
 RN [16]
 RP VARIANT XSCID ARG-115.
 RX MEDLINE-97042245; PubMed-8900089;
 RA Stephan V., Wahn V., Le Delst F., Dirksen U., Broeker B.,
 RA Mueller-Fleckenstein I., Horneff G., Schroten H., Fischer A.,
 RA de Saint Basile G.;
 RT "Atypical X-linked severe combined immunodeficiency due to possible
 RT spontaneous reversion of the genetic defect in T cells.";
 RL New Engl. J. Med. 335:1563-1567(1996).
 RN [17]
 RP VARIANT XSCID GLN-285.
 RX MEDLINE-97295088; PubMed-9150740;
 RA Jones A.M., Clark P.A., Katz F., Genet S., McMahon C., Alterman L.,
 RA Cant A., Kinon C.;
 RT "B-cell-negative severe combined immunodeficiency associated with a
 RT common gamma chain mutation.";
 RL Hum. Genet. 99:677-680(1997).
 RN [18]
 RP VARIANT XSCID CYS-222.
 RX MEDLINE-98064061; PubMed-9399950;
 RA Sharfe N., Shahar M., Rolfman C.M.;
 RT "An interleukin-2 receptor gamma chain mutation with normal thymus
 RT morphology.";
 RL J. Clin. Invest. 100:3036-3043(1997).
 CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
 CC INTERLEUKINS.
 CC -1- SUBUNIT: THE IL-13 RECEPTORS.
 CC PROBABLY ALSO THE IL-4, IL-7 AND
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A SEVERE COMBINED
 CC IMMUNODEFICIENCY, WHICH IS KNOWN AS AGAMMAGLOBULINEMIA, SWISS TYPE
 CC OR X-LINKED SEVERE COMBINED IMMUNODEFICIENCY DISEASE (XSCID).
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 CC -1- DATABASE: NAME=PROT: NOTE=X-CD guide CD132 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prot/cd/cd132.htm".
 CC -1- DATABASE: NAME=IL2Gbase; NOTE=X-linked SCID mutation database;
 CC WWW="http://www.hbgrl.nih.gov/DIR/GMBB/SCID/".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D11086; BAA01857.1; -;
 DR EMBL: L12183; AAA59145.1; -;
 DR EMBL: L12178; AAA59145.1; JOINED.
 DR EMBL: L12176; AAA59145.1; JOINED.
 DR EMBL: L12177; AAA59145.1; JOINED.
 DR EMBL: L12179; AAA59145.1; JOINED.
 DR EMBL: L12180; AAA59145.1; JOINED.
 DR EMBL: L12181; AAA59145.1; JOINED.
 DR EMBL: L12182; AAA59145.1; JOINED.
 DR EMBL: L19546; AAC37524.1; -;
 DR PIR: A42565; A42565;
 DR PDB: 1ILM; 26-JAN-95.
 DR PDB: 1ILN; 26-JAN-95.
 DR MIM: 308380; -;
 DR MIM: 300400; -;
 DR InterPro: IPR002996; CRLA.
 DR InterPro: IPR003961; FM_III.
 DR InterPro: IPR003531; Hematopo_receptor_S_F1.

Query Match 100.0%; Score 34; DB 1; Length 369;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
 |||||
 Db 181 HCLEH 185

```

RESULT 2
CYRG_CANFA STANDARD; PRT; 373 AA.
AC P40321;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cytokine receptor common gamma chain precursor (gamma-C)
DE (interleukin-2 receptor gamma chain) (IL-2R gamma chain) (p64).
GN IL2RG.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130114; PubMed=7829104;
RA Henthorn P.S., Somberg R.L., Fimlant V.M., Puck J.M., Patterson D.F.,
RA Felsburg P.J.;
RT "IL-2R gamma gene microdeletion demonstrates that canine X-linked
RT severe combined immunodeficiency is a homologue of the human
RT disease."
RL Genomics 23:69-74(1994).
CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
CC INTERLEUKINS.
CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC PROBABLY ALSO THE IL-13 RECEPTORS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A CANINE X-LINKED
CC SEVERE COMBINED IMMUNODEFICIENCY.
CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U04361; AAC48403.1; -.
CC HSSP: P31785; 11LN.
CC InterPro: IPR002996; CRIA.
CC InterPro: IPR003961; FN_III.
CC InterPro: IPR003531; Hematopo_receptor_S_F1.
CC Pfam: PF00041; fn3; 1.
CC SMART: SM00060; FN3; 1.
CC PROSITE: PS01355; HEMATOPO_REC_S_F1; 1.
CC Receptor; Transmembrane; Glycoprotein; Signal.
CC SIGNAL 1 22
CC CHAIN 1 22
CC POTENTIAL.
CC CYTOKINE RECEPTOR COMMON GAMMA CHAIN.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC FIBRONECTIN TYPE-III.
CC DOMAIN 284 373
CC TRANSSEM 262 283
CC DOMAIN 151 249
CC DISULFID 62 72
CC POTENTIAL.
CC DISULFID 102 115
CC POTENTIAL.
CC CARBOHYD 24 24
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 71 71
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 75 75
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 84 84
CC N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 373 AA; 42516 MW; 03A0DE1F8B089DBB CRC64;

Query Match 100.0%; Score 34; DB 1; Length 373;
 Best Local Similarity 100.0%; Pred. No. 8.9;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
 |||||
 Db 181 HCLEH 185

```

RESULT 3
CYRG_BOVIN STANDARD; PRT; 379 AA.
AC O95118;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cytokine receptor common gamma chain precursor (gamma-C)
DE (interleukin-2 receptor gamma chain) (IL-2R gamma chain) (p64).
GN IL2RG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96268473; PubMed=8672241;
RA Yeo J., Stone R.T., Solinas-Toldo S., Fries R., Beattie C.W.;
RT "Cloning and chromosomal mapping of bovine interleukin-2 receptor
RT gamma gene."
RL DNA Cell Biol. 15:453-459(1996).
CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
CC INTERLEUKINS.
CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC PROBABLY ALSO THE IL-13 RECEPTORS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U33748; AAB07812.1; -.
CC HSSP: P31785; 11LN.
CC InterPro: IPR002996; CRIA.
CC InterPro: IPR003961; FN_III.
CC InterPro: IPR003531; Hematopo_receptor_S_F1.
CC Pfam: PF00041; fn3; 1.
CC SMART: SM00060; FN3; 1.
CC PROSITE: PS01355; HEMATOPO_REC_S_F1; 1.
CC Receptor; Transmembrane; Glycoprotein; Signal.
CC SIGNAL 1 22
CC CHAIN 1 22
CC POTENTIAL.
CC CYTOKINE RECEPTOR COMMON GAMMA CHAIN.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC FIBRONECTIN TYPE-III.
CC DOMAIN 270 290
CC TRANSSEM 291 379
CC DOMAIN 158 256
CC DISULFID 68 78
CC POTENTIAL.
CC DISULFID 109 122
CC POTENTIAL.
CC CARBOHYD 77 77
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 81 81
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 90 90
CC N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 379 AA; 43037 MW; 33CFAD9C9B032178 CRC64;

Query Match 100.0%; Score 34; DB 1; Length 379;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HOLEH 5
 11111
 DB 188 HOLEH 192

RESULT 4
 GT29_FASHE STANDARD; PRT; 220 AA.
 ID GT29_FASHE
 AC P56598;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glutathione S-transferase 26 kDa 1 (EC 2.5.1.18) (GST1) (FHL) (GST
 DE class-alpha).
 OS Fasciola hepatica (Liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
 OC Rhabditiophora; Eulicthiophora; Revertospermatia; Mediofusata;
 OC Neodermata; Trematoda; Digenea; Echinostomata; Echinostomata;
 OC Fasciolidae; Fasciolidae; Fasciola.
 NC NCB1_TaxID=6192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92135306; PubMed=1740183;
 RA Panaccio M., Wilson L.R., Crameri S.L., Wajffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RL transerasas of Fasciola hepatica."
 RL Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Crameri S.L., Wajffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-386(1993).
 RN [3]
 RP SEQUENCE OF 22-220 FROM N.A.
 RA Crameri S.;
 RL Patent number WO9008819, 09-AUG-1990.
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: A00993; CA00118.1; -;
 DR HSSP: P31670; IPHE.
 DR InterPro: IPR004046; GST_C.
 DR InterPro: IPR004045; GST_N.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferase; Antigen; Multigene family.
 FT INT MET 0 BY SIMILARITY.
 FT CONFLICT 22 22 Y->V (IN REF. 3).
 FT CONFLICT 110 111 DP->VS (IN REF. 3).

FT CONFLICT 189 189 A->P (IN REF. 3).
 SQ SEQUENCE 220 AA; 25598 MW; 27B9F150B75D101F CRC64;

Query Match 91.2%; Score 31; DB 1; Length 220;
 Best Local Similarity 80.0%; Pred. No. 20;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HOLEH 5
 11111
 DB 166 HOLEH 170

RESULT 5
 PGK_CHICK STANDARD; PRT; 416 AA.
 ID PGK_CHICK
 AC P51903;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Phosphoglycerate kinase (EC 2.7.2.3).
 GN PGK.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae;
 OC Gallus.
 NC NCB1_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skletal muscle;
 RX MEDLINE=94238148; PubMed=8182283;
 RA Rauen K.A., Le Ciel C.D., Abbott U.K., Hutchison N.J.;
 RT "Localization of the chicken PGK gene to chromosome 4p by
 RL fluorescence in situ hybridization."
 RL J. Hered. 85:147-150(1994).
 CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate -> ADP + 3-
 CC phospho-D-glyceroyl phosphate.
 CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L37101; AAC42219.1; -;
 DR HSSP: P00560; 10PG.
 DR InterPro: IPR001576; PGK.
 DR Pfam: PF00162; PGK; 1.
 DR PRINTS: PR00477; PHGLYCKINASE.
 DR PROSITE: PS00111; PGLYCERATE_KINASE; 1.
 KW Transferase; Kinase; Glycolysis.
 FT INT MET 0 BY SIMILARITY.
 SQ SEQUENCE 416 AA; 44585 MW; 216A7A295590C31A CRC64;

Query Match 91.2%; Score 31; DB 1; Length 416;
 Best Local Similarity 80.0%; Pred. No. 37;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HOLEH 5
 11111
 DB 48 HOLEH 52

RESULT 6
 ADP_MOUSE STANDARD; PRT; 425 AA.
 ID ADP_MOUSE
 AC P43883;

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Adipophilin (Adipose differentiation-related protein) (ADRP).
GN ADRP OR ADRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adipocyte;
RX MEDLINE=92390349; PubMed=1518805;
RA Jiang H.P., Serrero G.;
RT "Isolation and characterization of a full-length cDNA coding for an
RT adipose differentiation-related protein";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7856-7860(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H; TISSUE=Adipose tissue;
RX MEDLINE=93315154; PubMed=8325636;
RA Eisinger D.P., Serrero G.;
RT "Structure of the gene encoding mouse adipose differentiation-related
RT protein (ADRP)";
RL Genomics 16:638-644(1993).
CC -1- FUNCTION: MAY BE INVOLVED IN DEVELOPMENT AND MAINTENANCE OF
CC ADIPOSE TISSUE.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- TISSUE SPECIFICITY: ADIPOSE TISSUE SPECIFIC. EXPRESSED ABUNDANTLY
CC AND PREFERENTIALLY IN FAT PADS.
CC -1- INDUCTION: BY DEXAMETHASONE.
CC -1- SIMILARITY: BELONGS TO THE PERILIPIN FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC CC EMBL; M93275; AAA37176.1; -;
DR EMBL; L09734; -; NOT_ANNOTATED_CDS.
DR MGD; MGI:87920; Adip.
DR InterPro: IPR004279; perilipin.
DR Pfam; PF03036; perilipin; 1.
KW Membrane.
SQ SEQUENCE 425 AA; 46664 MW; 82624E6CE3429C22 CRC64;
Query Match 91.2%; Score 31; DB 1; Length 425;
Best Local Similarity 80.0%; Pred. No. 38;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HCLEH 5
DB 299 HCVH 303
RESULT 7
GUNA_CLOTM STANDARD; PRT; 477 AA.
AC P04955;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (EGA) (Endo-1,4-beta-glucanase)
DE (Cellulase A).
GN CELA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxId=1515;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=85157393; PubMed=3980433;
RA Begun P., Cornet P., Aubert J.-P.;
RT "Sequence of a cellulase gene of the thermophilic bacterium
RT Clostridium thermocellum";
RL J. Bacteriol. 162:102-105(1985).
RN [2]
RP SEQUENCE OF 33-46.
RC STRAIN=JW20;
RX MEDLINE=96267082; PubMed=8664281;
RA Choi S.K., Ljungdahl L.G.;
RT "Dissociation of the cellulose of Clostridium thermocellum in the
RT presence of ethylenediaminetetraacetic acid occurs with the formation
RT of truncated polypeptides";
RL Biochemistry 35:4897-4905(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RX MEDLINE=96398613; PubMed=8805535;
RA Alzari P.M., Souchon H., Dominguez R.;
RT "The crystal structure of endoglucanase CelA, a family 8 glycosyl
RT hydrolase from Clostridium thermocellum";
RL Structure 4:265-275(1996).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC CC EMBL; K03088; AAA83521.1; -;
DR PIR; A23100; C2CLAM.
DR PDB; ICEM; 1J-JAN-97.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR002037; Glyco_hydro_8.
DR Pfam; PF00404; Dockerin_1; 2.
DR Pfam; PF01270; Glyco_hydro_8; 1.
DR PRINTS; PR00735; GLHYDRASE8.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
DR PROSITE; PS00448; CLOS_CELULOSE_RPT; 2.
DR PROSITE; PS00812; GLYCOSYL_HYDROL_F8; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal;
KW 3D-structure.
FT SIGNAL 1 32
FT CHAIN 33 477
FT ACT_SITE 95 95
FT ACT_SITE 152 152
FT DOMAIN 417 472
FT DOMAIN 417 440
FT REPEAT 449 472
FT REPEAT 41 41
FT CONFLICT 41 41
SQ SEQUENCE 477 AA; 52594 MW; BA5A0AD5022E8A51 CRC64;
Query Match 91.2%; Score 31; DB 1; Length 477;
Best Local Similarity 80.0%; Pred. No. 42;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HCLEH 5

```

Db      190 HCEVH 194

RESULT 8
ID      Y197_HUMAN      STANDARD:      PRT: 1314 AA.
AC      Q12769;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Protein KIAA0197 (Fragment).
GN      KIAA0197.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Bone marrow;
RA      MEDLINE=96281124; PubMed=8724849;
RA      Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT      "Prediction of the coding sequences of unidentified human genes. V.
RT      The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT      analysis of cDNA clones from human cell line KG-1.";
RL      DNA Res. 3:17-24(1996).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL: D83781; BAA12110.1; -
FT      NON-TER
FT      SEQUENCE 1314 AA; 148828 MW; B2FF622FF6A6AA0 CRC64;

Query Match      91.2%; Score 31; DB 1; Length 1314;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      1 HCLEH 5
Db      255 HCEVH 259

RESULT 9
ID      Y197_MOUSE      STANDARD:      PRT: 1402 AA.
AC      Q920W3;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Protein KIAA0197 (GTL-13).
GN      KIAA0197 OR GTL-13.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=129/SVJ;
RA      Van de Putte T., Cozijnsen M., Dewulf N., Tylianowski P., Lonnoy O.,
RA      Huylebroeck D.;
RT      "Mus musculus mRNA for gtl-13 (gene trap locus-13), similar to human
RT      KIAA0197 gene (D83781), complete cds.";
RL      Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL: AF104415; AAD17922.2; -
DR      MGD; MGI:1926227; Gtl1-13.
SQ      SEQUENCE 1402 AA; 158230 MW; 3BF5D9F057D28772 CRC64;

Query Match      91.2%; Score 31; DB 1; Length 1402;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      1 HCLEH 5
Db      257 HCEVH 261

RESULT 10
ID      AFAM_MOUSE      STANDARD:      PRT: 611 AA.
AC      O89020;
DT      15-JUL-1999 (Rel. 38, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Afamin precursor (Alpha-albumin) (Alpha-Alb).
GN      AFM.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=diaphragm;
RA      van Reeth T., Gabant P., Dreze P., Szpirer J., Szpirer C.;
RL      Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC      -----
CC      FUNCTION: POSSIBLE ROLE IN THE TRANSPORT OF YET UNKNOWN LIGAND.
CC      - SUBCELLULAR LOCATION: Extracellular.
CC      - DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC      - SIMILARITY: BELONGS TO THE ALB/AFM/DB FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL: AJ011080; CA009471.1; -
DR      HSSP; P02768; 18U5.
DR      InterPro: IPR000264; Serum_albumin.
DR      Pfam: PF00273; transport_prot. 3.
DR      PRINTS: PR00802; SERUMALBUMIN.
DR      ProDom: PD002486; Serum_albumin; 1.
DR      SMART: SM00103; ALBUMIN; 3.
DR      PROSITE: PS00212; ALBUMIN; 2.
KW      Transport; Albumin; Repeat; Glycoprotein; Signal.
FT      SIGNAL 1 21
FT      CHAIN 22 611 AFAMIN.
FT      REPEAT 22 210 1.
FT      REPEAT 211 402 2.
FT      REPEAT 403 611 3.
FT      DISULFID 77 86 BY SIMILARITY.
FT      DISULFID 99 114 BY SIMILARITY.
FT      DISULFID 113 124 BY SIMILARITY.
FT      DISULFID 148 193 BY SIMILARITY.
FT      DISULFID 224 270 BY SIMILARITY.
FT      DISULFID 269 277 BY SIMILARITY.
FT      DISULFID 289 303 BY SIMILARITY.
FT      DISULFID 302 313 BY SIMILARITY.

```


FT DISULFID 340 385 BY SIMILARITY.
 FT DISULFID 384 393 BY SIMILARITY.
 FT DISULFID 416 462 BY SIMILARITY.
 FT DISULFID 461 470 BY SIMILARITY.
 FT DISULFID 483 499 BY SIMILARITY.
 FT DISULFID 498 509 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT CARBOHYD 33 33 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 611 AA; 69635 MW; 39E46B6E723F89C8 CRC64;

Query Match 88.2%; Score 30; DB 1; Length 611;
 Best Local Similarity 80.0%; Pred. No. 83;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HCLEH 5
 Db 508 HCFEH 512

RESULT 11
 COTD_BACSU STANDARD; PRT; 75 AA.
 AC P07791;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Spore coat protein D.
 GN COTD.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88011308; PubMed=2821284;
 RA Donovan W., Zheng L., Sandman K., Losick R.;
 RT "Genes encoding spore coat polypeptides from Bacillus subtilis."
 RL J. Mol. Biol. 196;1-10(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / MARBURG;
 RX MEDLINE=96349105; PubMed=8760912;
 RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
 RA Serfor P.;
 RT "Sequence analysis of the Bacillus subtilis chromosome region between
 the sera and kgd loci cloned in a yeast artificial chromosome."
 RL Microbiology 142:2005-2016(1996).
 RN [3]
 RP SEQUENCE OF 1-6 FROM N.A.
 RX MEDLINE=90230303; PubMed=1691789;
 RA Zheng L., Losick R.;
 RT "Cascade regulation of spore coat gene expression in Bacillus
 subtilis."
 RL J. Mol. Biol. 212:645-660(1990).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X05681; CAA29168.1; ALT_TERM.
 DR EMBL; I47838; AAB38470.1; -;
 DR EMBL; Z99115; CAB14137.1; -;
 DR PIR; D27393; D27393.
 DR Subtilisin; Bg10493; cotd.

KW Sporulation; Complete proteome.
 SQ SEQUENCE 75 AA; 8840 MW; A5019889CA6CC0EA CRC64;

Query Match 85.3%; Score 29; DB 1; Length 75;
 Best Local Similarity 80.0%; Pred. No. 18;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HCLEH 5
 Db 17 HCFEH 21

RESULT 12
 CDAL_YEAST STANDARD; PRT; 312 AA.
 AC Q06703;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Chitin deacetylase 1 precursor (EC 3.5.1.41).
 GN CDAL OR YLR308W OR L2142.1.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 RA Maridis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
 RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,
 RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=97279228; PubMed=9133736;
 RA Mishra C.A., Sembo C.E., McCreath K.J., de la Vega H., Jones B.J.,
 RA Specht C.A., Robbins P.W.;
 RT "Cloning and expression of two chitin deacetylase genes of
 Saccharomyces cerevisiae."
 RL Yeast 13:327-336(1997).
 RL -1- FUNCTION: HYDROLYZES THE N-ACETAMIDO GROUPS OF N-ACETYL-D-
 GLUCOSAMINE RESIDUES IN CHITIN.
 CC -1- CATALYTIC ACTIVITY: Chitin + H(2)O = chitosan + acetate.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY DURING SPOULATION.
 CC -1- SIMILARITY: TO OTHER POLYSACCHARIDE DEACETYLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U17247; AAB67355.1; -;
 DR SGD; S0004298; CDAL.
 DR InterPro; IPR002509; Polysac_deacet.
 DR Pfam; PF01522; Polysac_deacet.1.
 KW Hydrolyase; Chitin degradation; Signal; Glycoprotein; Sporulation.
 FT SIGNAL 1 25 POTENTIAL
 FT CHAIN 26 312 CHITIN DEACETYLASE 1.
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 312 AA; 35693 MW; 4033EEFA803F6AF1 CRC64;

```

Query Match      85.3%; Score 29; DB 1; Length 312;
Best Local Similarity 80.0%; Pred. No. 68;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HCLSH 5
Db 107 HCTEH 111

RESULT 13
MHPB_ECOLI
ID MHPB_ECOLI STANDARD; PRT; 314 AA.
AC P54711: P77461; P77048;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 2,3-dihydroxyphenylpropionate 1,2-dioxygenase (EC 1.13.11.-).
GN MHPB OR B0348 OR Z0446 OR ECS0403.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RX MEDLINE=96359381; PubMed=8752345;
RA Spence E.L., Kawamukai M., Sanvoisin J., Braven H., Bugl T.D.H.;
RT "Catechol dioxygenases from Escherichia coli (Mhp) and Alcaligenes
RT eutrophus (Mpci): sequence analysis and biochemical properties of a
RT third family of extradiol dioxygenases.";
RL J. Bacteriol. 178:5249-5256(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / CS520;
RA Ferrandez A., Garcia J.L., Diaz E.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074933; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RTMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

```

```

RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kunara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE OF 1-15, AND CHARACTERIZATION.
RX MEDLINE=94002258; PubMed=8393388;
RA Bugl T.D.H.;
RT "Overproduction, purification and properties of 2,3-
RT dihydroxyphenylpropionate 1,2-dioxygenase from Escherichia coli.";
RL Biochim. Biophys. Acta 1207:258-264(1993).
CC -1- FUNCTION: EXTRADIOL CLEAVAGE OF 2,3-DIHYDROXYPHENYLPROPIONIC ACID.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: 3-HYDROXYPHENYLPROPIONATE DEGRADATION PATHWAY.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: TO A.EUTROPHUS MCP1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D86239; BAA13053.1; -
DR EMBL: Y09555; CAAT70748.1; -
DR EMBL: AE000142; AAC73451.1; -
DR EMBL: U73857; AAB18072.1; -
DR EMBL: AE005214; AAC54659.1; -
DR EMBL: AP002551; BAB33826.1; -
DR EcoGene: EG20274; mhpB.
DR InterPro: IPR004183; L19B.
DR Pfam: PF02900; L19B.
KW Oxidoreductase; dioxygenase; Aromatic hydrocarbons catabolism; Iron;
KW Repeat: Complete proteome.
FT DOMAIN 1 171 1.
FT DOMAIN 2 172 314 2.
FT CONFLICT 138 140 ING -> NKA (IN REF. 1).
FT CONFLICT 152 152 R -> H (IN REF. 1).
FT CONFLICT 157 157 A -> T (IN REF. 1).
SQ SEQUENCE 314 AA; 34196 MW; ELD5A857AE5DFE05 CMC64;

Query Match      85.3%; Score 29; DB 1; Length 314;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HCLSH 5
Db 6 HCLSH 10

RESULT 14
INHS_ECOLI
ID INHS_ECOLI STANDARD; PRT; 326 AA.
AC P76071;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase insH for insertion sequence element ISSY.
GN INSH OR B1370.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

```

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 MAU B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -1- FUNCTION: INVOLVED IN THE TRANSPOSITION OF THE INSERTION
 SEQUENCE IS5.
 CC -1- SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 15.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: AE000234; AAC74452.1; -;
 DR EcoGene: EG40008; insH.
 DR InterPro: IPR002559; Transposase_11.
 DR Pfam: PF01609; Transposase_11; 1.
 KW Transposable element; Transposition; DNA-binding; DNA recombination;
 KW Complete proteome.
 SQ SEQUENCE 326 AA; 37777 MW; 1A3C80D7FD5D8B5B CRC64;

Query Match 85.3%; Score 29; DB 1; Length 326;
 Best Local Similarity 60.0%; Pred. No. 71;
 Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HCLEH 5
 ||:|
 Db 64 HCMOH 68

RESULT 15
 HRP1_PLAFA STANDARD; PRT; 332 AA.
 ID HRP1_PLAFA
 AC P05227;
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 13-DEC-1998 (Rel. 37, Last annotation update)
 DE Histidine-rich protein precursor (Clone PHRP-II).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=86287395; PubMed=3016741;
 RX Wellens T.E., Howard R.J.;
 RT "Homologous genes encode two distinct histidine-rich proteins in a
 RT cloned isolate of Plasmodium falciparum.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6065-6069(1986).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: M13986; AA51639.1; -;
 DR EMBL: K03509; AA51639.1; JOINED.
 KW Malaria; Repeat; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 332 HISTIDINE-RICH PROTEIN.
 SQ SEQUENCE 332 AA; 35128 MW; 330842B99C5FCF5B CRC64;

Query Match 85.3%; Score 29; DB 1; Length 332;
 Best Local Similarity 80.0%; Pred. No. 73;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HCLEH 5
 ||:|
 Db 328 HCLRH 332

Search completed: August 20, 2002, 11:33:03
 Job time: 1452 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:31:56 ; Search time 191.14 Seconds
(without alignments)
4.525 Million cell updates/sec

Title: US-09-824-286-16
Perfect score: 34
Sequence: 1 HCLEH 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	172	16	09CM60
2	34	100.0	287	16	097QT4
3	34	100.0	314	10	09SD28
4	34	100.0	331	10	09SL79
5	34	100.0	332	10	09L80
6	34	100.0	351	10	09SIS5
7	34	100.0	364	10	09L0M0
8	34	100.0	375	13	09PDC9
9	34	100.0	701	4	096MA4
10	34	100.0	891	2	09JL38
11	34	100.0	1044	4	09Y4E4
12	34	100.0	1704	4	096L70
13	34	100.0	2117	2	09Z4Z9
14	32	94.1	85	16	09PBP3
15	32	94.1	294	2	09JRN7
16	32	94.1	317	13	09DGR3

17	32	94.1	364	5	09VDM7	09vdm7 drosophila
18	32	94.1	459	10	09FVK3	09fvk3 glycine max
19	32	94.1	1829	5	019815	019815 caenorhabd1
20	31	91.2	30	4	09P2Z8	09p2z8 homo sapien
21	31	91.2	103	12	09Y7G4	09y7g4 bovine aden
22	31	91.2	161	17	09YEG4	09yeg4 aetopyrum p
23	31	91.2	174	2	09AD00	09ad00 streptomyce
24	31	91.2	191	5	045682	045682 caenorhabd1
25	31	91.2	228	16	09CK00	09ck00 pasteurella
26	31	91.2	252	16	0984F0	0984f0 rhizobium 1
27	31	91.2	257	2	093SN7	093sn7 prochloroco
28	31	91.2	304	10	09SIS6	09sis6 arabidopsis
29	31	91.2	333	10	09FRR2	09frr2 arabidopsis
30	31	91.2	338	10	09SCX8	09scx8 arabidopsis
31	31	91.2	425	11	09CY19	09cy19 mus musculu
32	31	91.2	515	10	09SAK7	09sak7 arabidopsis
33	31	91.2	607	10	09PEP9	09pep9 arabidopsis
34	31	91.2	632	10	082261	082261 streptomyce
35	31	91.2	633	2	09F392	09f392 streptomyce
36	31	91.2	748	11	099MW0	099mw0 mus musculu
37	31	91.2	827	5	09U6M1	09u6m1 leishmania
38	31	91.2	1497	5	095YH6	095yh6 drosophila
39	31	91.2	1571	5	093YH7	093yh7 drosophila
40	31	91.2	1612	5	09VE37	09ve37 drosophila
41	31	91.2	2078	5	P91834	P91834 caenorhabd1
42	31	91.2	2454	3	09UVP2	09uvp2 emericella
43	31	91.2	2454	3	09UV56	09uv56 emericella
44	30	88.2	168	12	091MP9	091mp9 lumpy skin
45	30	88.2	175	2	044138	044138 anabena sp

ALIGNMENTS

RESULT 1
ID 09CM60 PRELIMINARY; PRT; 172 AA.
AC 09CM60;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE NPFA.
GN NPFA OR PM0981.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=2145866; PubMed=11248100;
RA May B.J., Zhang O., Li L.L., Paustian M.L., Whittram T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AEO06137; AAK03065.1; -;
DR InterPro; IPR000086; NUDIX_hydrolase.
DR PROSITE; PS00893; NUDIX; 1.
KW Complete proteome.
SQ SEQUENCE 172 AA; 20168 MW; F8FCB0B3E0B71F53 CRC64;

Query Match 100.0%; Score 34; DB 16; Length 172;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
Db 93 HCLEH 97
RESULT 2
Q97QT4 PRELIMINARY; PRT; 287 AA.

AC 097074;
DT 01-OCT-2001 (TREMblrel. 18, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR MUTR, PUTATIVE.
GN SP115.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RA MEDLINE=21357209; PubMed=11463916;
RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouiri H., Wolf A.M., Uitterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001)
DR EMBL: AE007412; AK75226.1; -
DR TIGR: SP115; -
DR InterPro: IPR001387; HTH_3.
DR InterPro: IPR003975; Shal_channel.
DR Pfam: PF01381; HTH_3; 1.
DR PRINTS: PR01497; SHALCHANNEL.
KW Complete proteome.
SQ SEQUENCE 287 AA; 34077 MW; B88A5C58FF6FE976 CRC64;

Query Match 100.0%; Score 34; DB 16; Length 287;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
Db 204 HCLEH 208

RESULT 3
Q9SDZ8 PRELIMINARY; PRT; 314 AA.
AC Q9SDZ8;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PUTATIVE PURPLE ACID PHOSPHATASE PRECURSOR.
GN PAP.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20314483; PubMed=10854785;
RA Schenk G., Guddat L.W., Ge Y., Carrington L.E., Hume D.A.,
RA Hamilton S., de Jersey J.;
RT "Identification of mammalian-like purple acid phosphatases in a wide
RT range of plants";
RL Gene 250:117-125(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=20517299; PubMed=11062342;
RA Oddie G.W., Schenk G., Angel N.Z., Walsh N., Guddat L.W.,
RA de Jersey J., Cassidy A.I., Hamilton S.E., Hume D.A.;
RT "Structure, function, and regulation of tartrate-resistant acid
RT phosphatase";

RL Bone 27:575-584(2000).
CC -I CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O -> AN
CC ALCOHOL + ORTHOPHOSPHATE.
DR EMBL: AF200827; AAF19823.1; -
DR InterPro: IPR003266; PA_phosphatase.
DR InterPro: IPR00934; Ser_thr_phosphatase.
DR Pfam: PF02227; PA_phosphatase; 1.
KW Hydrolyase; Iron; Signal; Zinc.
FT SIGNAL 1 30
FT CHAIN 31 314 POTENTIAL.
FT PUTATIVE PURPLE ACID PHOSPHATASE.
SQ SEQUENCE 314 AA; 35933 MW; 6982EBB4DEAC03CF CRC64;

Query Match 100.0%; Score 34; DB 10; Length 314;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
Db 254 HCLEH 258

RESULT 4
Q9LL79 PRELIMINARY; PRT; 331 AA.
AC Q9LL79;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PUTATIVE PURPLE ACID PHOSPHATASE.
GN PAP.
OS Phaseolus vulgaris (kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. RUFUS;
RX MEDLINE=20314483; PubMed=10854785;
RA Schenk G., Guddat L.W., Ge Y., Carrington L.E., Hume D.A.,
RA Hamilton S., de Jersey J.;
RT "Identification of mammalian-like purple acid phosphatases in a wide
RT range of plants";
RL Gene 250:117-125(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. RUFUS;
RX MEDLINE=20517299; PubMed=11062342;
RA Oddie G.W., Schenk G., Angel N.Z., Walsh N., Guddat L.W.,
RA de Jersey J., Cassidy A.I., Hamilton S.E., Hume D.A.;
RT "Structure, function, and regulation of tartrate-resistant acid
RT phosphatase";
RL Bone 27:575-584(2000).
DR EMBL: AF236109; AAF60317.1; -
DR InterPro: IPR005058; Peptidase_S26.
DR InterPro: IPR00934; Ser_thr_phosphatase.
DR PROSITE: PS00761; SPASE_I_3; UNKNOWN_1.
SQ SEQUENCE 331 AA; 37652 MW; 662A8A2D9EF3788 CRC64;

Query Match 100.0%; Score 34; DB 10; Length 331;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
Db 244 HCLEH 248

RESULT 5
Q9LL80 PRELIMINARY; PRT; 332 AA.

AC 09L180;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE PUTATIVE PURPLE ACID PHOSPHATASE.
 GN PAP.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 OX NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20314483; PubMed=10854785;
 RA Schenk G., Guddat L.W., Ge Y., Carrington L.E., Hume D.A.,
 RA Hamilton S., de Jersey J.;
 RT "Identification of mammalian-like purple acid phosphatases in a wide
 RT range of plants."
 RL Gene 250:117-125(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20517299; PubMed=11062342;
 RA Oddie G.W., Schenk G., Angel N.Z., Walsh N., Guddat L.W.,
 RA de Jersey J., Cassidy A.I., Hamilton S.E., Hume D.A.;
 RT "Structure, function, and regulation of tartrate-resistant acid
 RT phosphatase."
 RL Bone 27:575-584(2000).
 DR EMBL: AF236108; AAF60316.1; -;
 DR InterPro: IPR000934; Ser_thr.phosphatse.
 SQ SEQUENCE 332 AA; 37764 MW; B9E1463A946D9B49 CRC64;

Query Match 100.0%; Score 34; DB 10; Length 332;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
 |||||
 Db 252 HCLEH 256

RESULT 6
 ID 09SIS5 PRELIMINARY; PRT; 351 AA.
 AC 09SIS5;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE PUTATIVE PURPLE ACID PHOSPHATASE.
 GN AT2G01890.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=CV. COLUMBIA;
 MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buel C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Coppenhaver G.P., Preuss D., Nierman W.C., Venter J.C.;
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana."
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O - AN
 CC ALCOHOL + ORTHOPHOSPHATE.
 DR EMBL: AC007069; AAD21785.1; -;
 DR InterPro: IPR003266; PA_phosphatase.
 DR Pfam: PF02227; PA_phosphatase; 1.
 DR HydroLase; Iron; Zinc.
 SQ SEQUENCE 351 AA; 39998 MW; 335136FA234E8916 CRC64;

Query Match 100.0%; Score 34; DB 10; Length 351;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
 |||||
 Db 270 HCLEH 274

RESULT 7
 ID 09LOW0 PRELIMINARY; PRT; 364 AA.
 AC 09LOW0;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE F10B6.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chao Q., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,
 RA Williams S., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
 RA Hansen N.F., Huizar L., Kremetska I., Lenz C., Li J., Liu S.,
 RA Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaya V., Yu G.,
 RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome
 RT 1."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chiu J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Theologis A., Ecker J.;
 RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC006917; AAF79221.1; -;
 DR InterPro: IPR000934; Ser_thr.phosphatse.
 SQ SEQUENCE 364 AA; 41839 MW; 192AA2367701CEC0 CRC64;

Query Match 100.0%; Score 34; DB 10; Length 364;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
 |||||
 Db 281 HCLEH 285

RESULT 8
Q9DDC9 PRELIMINARY; PRT; 375 AA.
ID Q9DDC9
AC Q9DDC9
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CHICKEN GONADOTROPIN-RELEASING HORMONE RECEPTOR.
GN CGNRH-R.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21269321; PubMed=11112780;
RA Sun Y.M., Flanagan C.A., Illing N., Ott T.R., Sellar R., Fromme B.J.,
Hapgood J., Sharp P., Sealton S.C., Millar R.P.;
RT "A chicken gonadotropin-releasing hormone receptor that confers
agonist activity to mammalian antagonists: identification of D-lys6 in
the ligand and extracellular loop two of the receptor as
determinants.";
RT J. Biol. Chem. 276:7754-7761(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Sun Y.M.;
RL Thesis (1998), Department of Chemical Pathology,
University of Cape Town, Cape Town, South Africa.
DR EMBL: AJ304414; CAC18674.1;-
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCR_HODOPSN.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN_1.
DR PROSITE: PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 375 AA; 42097 MW; 7DD866F0C6B5CE4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 34; DB 13; Length 375;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 HCLEH 5
Db 28 HCLEH 32
RESULT 9
Q96MA4 PRELIMINARY; PRT; 701 AA.
ID Q96MA4
AC Q96MA4
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CDNA FLJ32706 FTS; CLONE TEST1200627 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
Holtu T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
Matanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;

RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK057268; BAB71404.1;-
FT NON_TER
SQ SEQUENCE 701 AA; 79822 MW; 99F46B2B8BD/BCA1 CRC64;

Query Match
Best Local Similarity 100.0%; Score 34; DB 4; Length 701;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 HCLEH 5
Db 517 HCLEH 521
RESULT 10
Q93L38 PRELIMINARY; PRT; 891 AA.
ID Q93L38
AC Q93L38
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MEMBRANE TRANSDUCATOR.
GN AMPT.
OS Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HHL;
RX MEDLINE=93209944; PubMed=8458843;
RA Ueda K., Miyake K., Horinouchi S., Beppu T.;
RT "A gene cluster involved in aerial mycelium formation in Streptomyces
griseus encodes proteins similar to the response regulators of Two-
component regulatory systems and membrane translocators.";
RT J. Bacteriol. 175:2006-2016(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HHL;
RX MEDLINE=98422460; PubMed=9748440;
RA Ueda K., Hsieh C.-W., Tosaki T., Shinkawa H., Beppu T., Horinouchi S.;
RT "Characterization of an A-factor-responsive repressor for amr
essential for onset of aerial mycelium formation in Streptomyces
griseus.";
RL J. Bacteriol. 180:5085-5093(1998).
DR EMBL: AB006206; BAB62264.1;-
SQ SEQUENCE 891 AA; 94161 MW; 7CF1487510BCC304 CRC64;

Query Match
Best Local Similarity 100.0%; Score 34; DB 2; Length 891;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 HCLEH 5
Db 107 HCLEH 111
RESULT 11
Q9Y4E4 PRELIMINARY; PRT; 1044 AA.
ID Q9Y4E4
AC Q9Y4E4
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE KIA00580 PROTEIN (FRAGMENT).
GN KIA00580.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=98290545; PubMed=628581;
 RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can RT code for large proteins in vitro."
 RL DNA Res. 5:31-39(1998).
 DR EMBL: AB011152; BAA25506.1; -
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR000198; RhogAP.
 DR InterPro: IPR001164; Znf_GCS.
 DR Pfam: PF01412; ArfGAP; 1.
 DR Pfam: PF00169; PH; 4.
 DR Pfam: PF00620; RhogAP; 1.
 DR PRINTS: PR00405; REVINTRACTING.
 DR SMART: SM00105; ArfGAP; 1.
 DR SMART: SM00233; PH; 3.
 DR SMART: SM00324; RhogAP; 1.
 DR PROSITE: PSS0003; PH_DOMAIN; 3.
 FT NON_TER 1
 SQ SEQUENCE 1044 AA; 119464 MW; 68E22A6D5EA53BD CRC64;

Query Match 100.0%; Score 34; DB 4; Length 1044;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
 |||||
 Db 954 HCLEH 958

RESULT 12
 ID 096L70 PRELIMINARY; PRT; 1704 AA.
 AC 096L70;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE ARA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Miura K., Jacques K.M., Stauffer S., Kubosaki A., Zhu K., Resau J., Zheng Y., Randazzo P.A.;
 RT "ARAP1, a point of convergence for phosphoinositide, Arf and Rho RT signaling, regulates both membrane and actin cytoskeleton remodeling."
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY049733; AAL12170.1; -
 SQ SEQUENCE 1704 AA; 193422 MW; 3770EDE08E42E2CC CRC64;

Query Match 100.0%; Score 34; DB 4; Length 1704;
 Best Local Similarity 100.0%; Pred. No. 16+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
 |||||
 Db 1614 HCLEH 1618

RESULT 13
 ID 092429 PRELIMINARY; PRT; 2117 AA.
 AC 092429;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CDA PEPTIDE SYNTHETASE III (FRAGMENT).
 GN SCE63.01.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redendach M., Kleiser H.M., Denaplatte D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL035707; CAB38876.1; -
 DR HSSP: P14687; IAMU.
 DR InterPro: IPR000873; AMP-bind.
 DR InterPro: IPR001242; DUF4.
 DR InterPro: IPR00379; Est_lip_thioest_actsite.
 DR InterPro: IPR003880; Phosphopant_attach.
 DR InterPro: IPR001031; Thioesterase.
 DR Pfam: PF00501; AMP-binding; 2.
 DR Pfam: PF00668; Condensation; 1.
 DR Pfam: PF00550; PP-binding; 2.
 DR Pfam: PF00975; Thioesterase; 1.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PSS0075; ACP_DOMAIN; 2.
 DR PROSITE: PSS00455; AMP-BINDING; 2.
 DR PROSITE: PSS00012; PHOSPHOPANTHETINE; UNKNOWN; 2.
 FT NON_TER 1
 SQ SEQUENCE 2117 AA; 225115 MW; 30EBF48688DBBEA1 CRC64;

Query Match 100.0%; Score 34; DB 2; Length 2117;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5
 |||||
 Db 466 HCLEH 470

RESULT 14
 ID 09BPB3 PRELIMINARY; PRT; 85 AA.
 AC 09BPB3;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN XF2097.
 GN XF2097.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,

RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber L.R.,
RA Ho P.L., Honzel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A., Jr., Nodrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.,
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
RL Nature 406:151-159(2000).
DR EMBL: AE004025; AAF64896.1; -
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 85 AA; 10186 MW; 361F2727BF73718F CRC64;

Query Match 94.1%; Score 32; DB 16; Length 85;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HCLEH 5
||:|
Db 15 HCLIEH 19

RESULT 15
Q9JRN7 PRELIMINARY; PRT; 294 AA.
AC Q9JRN7;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE HYPOTHETICAL 33.3 KDA PROTEIN.
OS Actinobacillus actinomycetemcomitans (Haemophilus
OS actinomycetemcomitans).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SUNYAB 75;
RA Suzuki N., Nakano Y., Yoshida Y., Nako H., Yamashita Y., Koga T.,
RT "Genetic analysis of the gene cluster for the synthesis of serotype a-
RT specific polysaccharide antigen in *A. actinomycetemcomitans*,"
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB046360; BAB03206.1; -
KW Hypothetical protein.
SQ SEQUENCE 294 AA; 33342 MW; ABC3F7E2C8CODEB4 CRC64;

Query Match 94.1%; Score 32; DB 2; Length 294;
Best Local Similarity 80.0%; Pred. No. 56;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5
||:|
Db 133 HCLMEH 137

Search completed: August 20, 2002, 11:31:58
Job time: 1472 sec

100

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:07:18 ; Search time 187.18 Seconds
(without alignments)
1.780 Million cell updates/sec

Title: US-09-824-286-17
Perfect score: 19
Sequence: 1 FNP 3

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_032802:*

- 1: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	3	19	AAW31654
2	19	100.0	4	12	AA13261
3	19	100.0	5	20	AA131285
4	19	100.0	6	15	AA135366
5	19	100.0	6	15	AA135569
6	19	100.0	6	19	AA1369443
7	19	100.0	6	20	AA138139
8	19	100.0	6	21	AA1395393
9	19	100.0	6	21	AA1384078
10	19	100.0	7	15	AA135138
11	19	100.0	7	16	AA1365309

12	19	100.0	7	19	AAW52004	Peptide having imm
13	19	100.0	7	19	AAW51081	Pentillanin cont
14	19	100.0	7	22	AAW44049	H11 binding site c
15	19	100.0	7	22	AAW44489	H11 binding site c
16	19	100.0	7	22	AAW46402	H11 binding site c
17	19	100.0	7	22	AAW46568	H11 binding site c
18	19	100.0	7	22	AAW46608	H11 binding site c
19	19	100.0	8	14	AAW32742	PK40 TAV/neurofila
20	19	100.0	8	16	AAW72720	MRSA-230 tryptic p
21	19	100.0	8	16	AAW73401	Human TSH receptor
22	19	100.0	8	16	AAW73402	Human TSH receptor
23	19	100.0	8	16	AAW73403	Human TSH receptor
24	19	100.0	8	19	AAW69539	Human 20K growth h
25	19	100.0	8	20	AAW27417	PK40 protein kinase
26	19	100.0	9	15	AAW38111	Hepatitis B virus-
27	19	100.0	9	15	AAW55144	Transglutaminase p
28	19	100.0	9	16	AAW5315	Paqrus major trans
29	19	100.0	9	18	AAW33943	Beta1-adrenergic r
30	19	100.0	9	18	AAW36467	Cyclitol-ubiquitin
31	19	100.0	9	19	AAW83244	NPF motif EH domai
32	19	100.0	9	20	AAW45679	Immunogenic peptid
33	19	100.0	9	20	AAW46662	Immunogenic peptid
34	19	100.0	9	20	AAW07211	Hydrophobic sequen
35	19	100.0	9	20	AAW98143	Transglutaminase c
36	19	100.0	9	21	AAW03485	Human tyrosine-rel
37	19	100.0	9	22	AAW03833	Hepatitis C virus
38	19	100.0	9	22	AAW50026	Mutant antimesothe
39	19	100.0	10	15	AAW38149	Hepatitis B virus-
40	19	100.0	10	18	AAW23359	Common receptor su
41	19	100.0	10	19	AAW69540	Human 20K growth h
42	19	100.0	10	19	AAW69541	Human 20K growth h
43	19	100.0	10	20	AAW45717	Immunogenic peptid
44	19	100.0	10	22	AAW99103	Potato patatin pro
45	19	100.0	10	22	AAW99104	Potato patatin pro

ALIGNMENTS

RESULT	1
AAW31654	
ID	AAW31654 standard; Peptide: 3 AA.
XX	
AC	AAW31654:
XX	
DT	21-MAY-1998 (first entry)
XX	
DE	Human cytokine receptor gc chain epitope.
XX	
KW	Cytokine receptor; gamma common chain; gc chain; human;
KW	blocking agent; monoclonal antibody; Cp.B8; immunological disease;
KW	myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
KW	insulin-dependent diabetes; inflammatory bowel disease;
KW	sympathetic ophthalmia; uveitis; allergy; asthma; infection;
KW	graft versus host disease; psoriasis; immunosuppressive; therapy;
XX	epitope.
OS	Homo sapiens.
XX	
PN	WO9743416-A1.
XX	
PD	20-NOV-1997.
XX	
PF	09-MAY-1997; 97WO-US07870.
XX	
PR	10-MAY-1996; 96US-0017466.
XX	
PA	(BIOJ) BIOGEN INC.
XX	
PI	Benjamin CD, Burkiy LC, Hession C, Whitty A;
XX	
DR	WPI; 1998-008885/01.
XX	

PT Blockin agents of the gamma common chain of cytokine receptors -
PT particularly monoclonal antibodies, used to induce T cell anergy for
PT treatment of immunological diseases
XX
XX
FS Claim 24; Page 85; 11pp; English.
XX
XX This peptide comprises an epitope of the human cytokine receptor
CC common gamma (gc) chain (see AAW31646) that is recognised by
CC gc blocking agents of the invention. 5 Such epitopes (see
CC AAW31650-54) have been identified. The invention provides
CC compositions and methods for inhibiting cytokine signalling using
CC gc chain blocking agents for the treatment of immunological
CC diseases such as myasthenia gravis, rheumatoid arthritis, lupus,
CC multiple sclerosis, insulin-dependent diabetes, inflammatory bowel
CC disease, sympathetic ophthalmia, uveitis, allergy, asthma,
CC parasitic infection, graft vs. host disease or psoriasis. A
CC preferred gc blocking agent is Mab CP.88 or its Fab fragment (see
CC also AAW31647-48).
XX
XX Sequence 3 AA:

Query Match	100.0%	Score 19;	DB 19;	Length 3;
Best Local Similarity	100.0%	Pred. NO. 6.4e+05;		
Matches	3;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;
QY	1 FNP 3			
db	1 fnp 3			

	RESULT	2
XX	AAR13261	
ID	AAR13261 standard; Protein; 4 AA.	
XX		
AC	AAR13261;	
XX		
DT	11-OCT-1991 (first entry)	
DE		
XX	Cytotoxic Cell Protease Inhibitor EF2372.	
KM		
XX	hCCP inhibitor; cytotoxic T-lymphocytes; CTL.	
OS	Homo sapiens.	
XX		
PN	WO9110685-A.	
XX		
PD	25-JUL-1991.	
XX		
PF	17-JAN-1991; 91WO-US00340.	
XX		
PR	19-JAN-1990; 90US-0467880.	
XX		
PA	(SERA-) SERAGEN INC.	
PI		
XX	Bleackley RC, Lobe CG, Paetkau VH, James MN, Murphy M:	
DR	WPI: 1991-237989/32.	
PT	DNA vectors, and inhibitors of cytotoxic cell protease - for	
PT	treatment of auto-immune diseases e.g. pernicious anaemia,	
PT	rheumatoid arthritis, allo-graft rejection etc.	
PS		
XS	Claim 9; Fig 11; 62pp; English.	
CC	CDNAs encoding murine and human CCP's were isolated and sequenced.	
CC	The amino acid sequences they encode were deduced and computer	
CC	analysis of the protein structure was performed. This protease	
CC	inhibitor is one of nine tetrapeptides of the invention which act	
CC	as competitive inhibitors. They are based on the protease cleavage	
CC	sites of CCP substrate molecules which were inferred from the	
CC	analysis of the protease binding pocket.	
CC	See AA012862-6 and AAR13254-R13362.	

```
xx      Sequence      4 AA;
SQ
Query Match          100.0%; Score 19; DB 12; Length 4;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 FNP 3
        |||
Db       1 fnp 3
```

RESULT	3
AAV31295	AAV31295 standard; peptide; 5 AA.
ID	AAV31295 standard; peptide; 5 AA.
AC	AAV31295;
XX	
DT	28-SEP-1999 (first entry)
XX	
DE	Beta-adrenergic GPCR consensus peptide motif.
XX	
KM	Gene characterisation; recombinase; homology clamp; homology motif tag;
KW	HMT; genetic manipulation; genetic engineering; gene transcription; GPCR;
KX	drug target; homologous recombination; target identification; TGF-beta;
XX	G-protein coupled receptor; transforming growth factor-beta.
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 1 /label= Ala or Gly
FT	Misc-difference 5 /label= Ile or Leu
XX	
PN	WO937755-A2.
PD	
PD	29-JUL-1999.
PE	11-DEC-1998; 98WO-US26498.
PR	11-DEC-1997; 97US-0070734.
PA	(PANG-) PANGENE CORP.
PI	Lehman CW, Patl S, Zarling D, Zeng H;
DR	WPI; 1999-458689/38.
PT	New compositions and methods for targeting sequence modifications in related family genes
PS	
XX	Disclosure; Fig 1B; 46pp; English.

The invention provides compositions and methods for the evaluation and characterisation of individual and sets of genes in disease states. The composition comprises at least one recombinase and at least two single-stranded targeting polynucleotides which are substantially complementary to each other and each having a consensus homology clamp for a gene family i.e. a homology motif tag (HMT). The composition is useful in kit form which include the composition as libraries or pools of degenerate cssDNA probes along with other reagents such as recombinase etc. The methods and compositions are used for inactivation of a gene family genome i.e. exogenous targeting polynucleotides can be used to inactivate, decrease or alter the biological activity of one or more genes in a cell (or transgenic nonhuman animal or plant). This is useful for generating animal models of disease, or in the elucidation of gene function and activity. Alternatively, the biological activity of the wild-type gene may be either decreased or the wild-type activity altered to mimic disease states. This includes genetic manipulation of non-coding gene sequences that affect the transcription of genes, including promoter, enhancers and transcriptional activating sequences. The

CC compositions are useful in identifying new members of gene families
 CC which may be useful in functional genomic studies as well as in
 CC identification of new drug targets. HMWs used in homologous recombination
 CC methods can generate animals that have a wide variety of mutations in a
 CC wide variety of related genes, potentially resulting in a wide variety of
 CC phenotypes including those related to disease states. This may also be
 CC done on a cellular level to identify genes involved in cellular
 CC phenotypes i.e. target identification. Sequences AAY31248-308 represent
 CC consensus peptide motifs of the G-protein coupled receptor (GPCR) gene
 CC family and transforming growth factor-beta (TGF-beta) gene family.
 XX
 SQ Sequence 5 AA;

Query Match 100.0%; Score 19; DB 20; Length 5;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
 |||
 Db 2 fnp 4

RESULT 4
 AAR55566
 ID AAR55566 standard; Protein: 6 AA.
 XX

AC AAR55566;

DT 10-FEB-1995 (first entry)

DE Epitope recognised by MAB 87-55/02/2, isolated from random library.

KM epitope; binding specificity: random hexapeptide; ligand screening;
 KW outer membrane protein; ompa; cytomagalovirus; CMV; ppl50 protein.

OS Synthetic.

PN EP603672-A.

PD 29-JUN-1994.

PF 10-DEC-1993; 93EP-0119921.

PR 23-DEC-1992; 92DE-4243770.

PA (BEHW) BEHRINGWERKE AG.

PI Grundmann U, Wissel T, Zetlmeissl G;

DR WPI; 1994-201750/25.

XX
 PT Prepn. of peptide(s) with specific binding properties - by
 PT expressing oligo-nucleotide(s) in bacteria or animal cells, as
 PT surface proteins, then screening with specific receptor attached
 PT to solid phase
 XX

PS Example 4; Page 10; 25pp; German.

CC A library of random hexapeptides was screened for epitopes which
 CC are recognised by the Monoclonal antibody MAB 87-55/02/2. The MAB
 CC is known to recognise an epitope of the ppl50 protein of
 CC cytomagalovirus. Antibody screening identified 5 hexapeptide motifs
 CC (AAR55566-R55570), all of which differed from the wild-type epitope.
 XX
 SQ Sequence 6 AA;

Query Match 100.0%; Score 19; DB 15; Length 6;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3

Db 1 fnp 3

RESULT 5
 AAR55569
 ID AAR55569 standard; Protein: 6 AA.
 XX

AC AAR55569;

DT 10-FEB-1995 (first entry)

DE Epitope recognised by MAB 87-55/02/2, isolated from random library.

KM epitope; binding specificity: random hexapeptide; ligand screening;
 KW outer membrane protein; ompa; cytomagalovirus; CMV; ppl50 protein.

OS Synthetic.

PN EP603672-A.

PD 29-JUN-1994.

PF 10-DEC-1993; 93EP-0119921.

PR 23-DEC-1992; 92DE-4243770.

PA (BEHW) BEHRINGWERKE AG.

PI Grundmann U, Wissel T, Zetlmeissl G;

DR WPI; 1994-201750/25.

XX
 PT Prepn. of peptide(s) with specific binding properties - by
 PT expressing oligo-nucleotide(s) in bacteria or animal cells, as
 PT surface proteins, then screening with specific receptor attached
 PT to solid phase
 XX

PS Example 4; Page 10; 25pp; German.

CC A library of random hexapeptides was screened for epitopes which
 CC are recognised by the Monoclonal antibody MAB 87-55/02/2. The MAB
 CC is known to recognise an epitope of the ppl50 protein of
 CC cytomagalovirus. Antibody screening identified 5 hexapeptide motifs
 CC (AAR55566-R55570), all of which differed from the wild-type epitope.
 XX

SQ Sequence 6 AA;

Query Match 100.0%; Score 19; DB 15; Length 6;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
 |||
 Db 1 fnp 3

RESULT 6
 AAM69443
 ID AAM69443 standard; Protein: 6 AA.
 XX

AC AAM69443;

DT 10-DEC-1998 (first entry)

DE Rat phospholipase A1 fragment.

KW Phospholipase A1; phosphatidyl-serine hydrolysis; serine phospholipid;
 KW PLA1; enzyme; rat.

OS Rattus sp.

XX

PN JP10201479-A.
XX
PD 04-AUG-1998.
XX
PE 23-JAN-1997; 97JP-0024269.
XX
PR 23-JAN-1997; 97JP-0024269.
XX
PA (TORA) TORAY IND INC.
XX
DR WPI: 1998-474493/41.
XX
PT Phospholipase A1 and nucleic acid coding for it - having substrate
PT specifically against serine phospholipid
XX
PS Example 1; Page 5; 13pp; Japanese.
XX
CC This sequence is a fragment of the rat phospholipase A1 (PLA1) of the
CC invention. PLA1 is capable of hydrolysing phosphatidyl-serine, and has
CC substrate specificity for serine phospholipids.
XX
SQ Sequence 6 AA:

Query Match 100.0%; Score 19; DB 19; Length 6;
Best Local Similarity 100.0%; Pred. No. 6,4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 FNP 3
|||
Db 1 fnp 3

RESULT 7
AAW98139
ID AAW98139 standard; Peptide: 6 AA.
XX
AC AAW98139;
XX
DT 05-JUL-1999 (first entry)
XX
DE Transglutaminase conserved peptide.
XX
KW Transglutaminase; diagnosis; therapy.
XX
OS Homo sapiens.
OS Mus sp.
OS Rattus sp.
XX
PN WO9910507-A1.
XX
PD 04-MAR-1999.
XX
PF 28-AUG-1998; 98WO-US17857.
XX
PR 29-AUG-1997; 97US-0920919.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Aeschlimann DP, Mosher DF;
XX
DE WPI: 1999-190622/16.
XX
PT Novel transglutaminase and new degenerate and specific primers -
PT useful in diagnostic assays for amplification and characterization
PT of transglutaminase genes and proteins
XX
PS Disclosure; Page 14; 68pp; English.
XX
CC The present sequence is a conserved region of human, mouse and
CC rat transglutaminases. Short (18-mer) oligonucleotides based
CC on this peptide were unsuccessful as PCR primers for amplifying
CC transglutaminase gene sequences. The invention provides

CC degenerate and specific primers (see AAX24958, AAX24960-73) for
CC the PCR amplification of known and novel transglutaminase
CC genes, including the novel human transglutaminase Tgx gene (see
CC AAX24956).
XX
SQ Sequence 6 AA:

Query Match 100.0%; Score 19; DB 20; Length 6;
Best Local Similarity 100.0%; Pred. No. 6,4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
|||
Db 2 fnp 4

RESULT 8
AAV95393
ID AAV95393 standard; Peptide: 6 AA.
XX
AC AAV95393;
XX
DT 25-SEP-2000 (first entry)
XX
DE Human pancreatic polypeptide zslg66 motif 2.
XX
KW zslg66; human; pancreas; infection; diabetes; diagnosis; therapy.
XX
OS Homo sapiens.
XX
PN WO200036104-A1.
XX
PD 22-JUN-2000.
XX
PF 14-DEC-1999; 99WO-US29669.
XX
PR 16-DEC-1998; 98US-0212947.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Sheppard PO;
XX
DR WPI: 2000-475395/41.
XX
PT Nucleic acids encoding pancreatic zslg66 proteins useful for
PT identifying compounds that may be used to treat, for example microbial
PT infections -
XX
PS Disclosure; Page 14; 108pp; English.
XX
CC The present sequence is that of peptide motif 2 of zslg66 (see
CC AAV95391), a novel human secreted pancreatic polypeptide. The motif
CC can be used to design oligonucleotide primers (see AAX9885) useful
CC for identifying related sequences. zslg66 polypeptides can be
CC used in the prevention, treatment and diagnosis of diseases
CC associated with inappropriate zslg66 expression, to raise
CC antibodies, and to screen for modulators (agonists and antagonists)
CC of zslg66 activity.
XX
SQ Sequence 6 AA:

Query Match 100.0%; Score 19; DB 21; Length 6;
Best Local Similarity 100.0%; Pred. No. 6,4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
|||
Db 3 fnp 5

RESULT 9

AAy84078
ID AAY84078 standard; Peptide; 6 AA.
XX
AC AAY84078;
XX
DT 03-JUL-2000 (first entry)
XX
DE Amino acid sequence of the C-terminal of a defensin fusion protein.
XX
KW Antimicrobial protein; AMP1; transgenic plant; linker propeptide;
XX protein expression; plant defensin; RsaFP2; antifungal protein; AFP2.
OS Synthetic.
XX
PN W0200011175-A1.
XX
PD 02-MAR-2000.
XX
PF 17-AUG-1999; 99WO-GB02716.
XX
PR 18-AUG-1998; 98GB-0018001.
XX 04-DEC-1998; 98GB-0026753.
XX
PA (ZENE) ZENECA LTD.
XX
PI Broekaert WF, Francois IEJA, De Bolle MFC, Evans IJ, Ray JA;
XX WPI; 2000-246564/21.
XX
XX Improving expression of polyproteins in plants involves coexpression of
PT two or more proteins in plants within a single transcription unit -
XX
XX Example 10; Page 40; 151pp; English.
PS
XX The present sequence represents the C-terminal of a fusion protein
CC of the invention, comprising the Dahlia antimicrobial protein (AMP) 1
CC and an antifungal protein (RsaFP2), linked by a linker propeptide. The
CC specification describes methods for improving expression levels of one
CC or more proteins in a transgenic plant. The method comprises inserting a
CC DNA sequence having a promoter region operably linked to two or more
CC protein encoding regions separated by a DNA sequence coding for a
CC linker propeptide and a terminator region. The method is used to
CC produce proteins in plants. The linker propeptide comprising a cleavage
CC site, whereby the expressed polypeptide is post-translationally
CC processed into the component protein molecules. The propeptide sequence
CC is rich in amino acids A, V, S and T and contains dipeptidic sequences
CC consisting of either two acidic, two basic or one acidic and one basic
CC residue as a cleavable linker sequence.
XX
SQ Sequence 6 AA:

Query Match 100.0%; Score 19; DB 21; Length 6;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FNP 3
|||
Db 3 fnp 5

RESULT 10
AAR5138
ID AAR5138 standard; peptide; 7 AA.
XX
AC AAR5138;
XX
DT 16-DEC-1994 (first entry)
XX
DE Transglutaminase peptide.
XX
KW Transglutaminase; TGase; fish; liver; Teragra chalcogramma;
KW Pagrus major; enzyme; acyl transfer; gamma-carboxyamide;
CC

KW glutamine.
XX
XX Teragra chalcogramma.
OS
XX JP06113844-A.
XX
XX
XX 26-APR-1994.
PD
XX
XX 15-DEC-1992; 92JP-0334224.
PF
XX 19-AUG-1992; 92JP-0220296.
PR
XX
XX (AJIN) AJINOMOTO KK.
PA
XX
XX WPI; 1994-172742/21.
DR
XX
XX Compns. contg. trans-glutaminase derived from fish liver - useful
PT for prepn. of protein-gelled prod., esp. food prod.
PT
XX
XX Disclosure; Page 10; 26pp; Japanese.
PS
XX
XX A compsn. contains at least 0.5 units per mg protein of a novel
CC transglutaminase (TGase). The TGase is derived from fish liver
CC (esp. Teragra chalcogramma and Pagrus major) and catalyses the
CC acyl transfer reaction between the gamma-carboxyamide gp. of the
CC glutamine residue of a peptide chain and various prim. amines.
CC TGase is useful for prepn. of protein-gelled prods., esp. food.
XX
XX
SQ Sequence 7 AA:

Query Match 100.0%; Score 19; DB 15; Length 7;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FNP 3
|||
Db 5 fnp 7

RESULT 11
AAR65309
ID AAR65309 standard; Peptide; 7 AA.
XX
AC AAR65309;
XX
DT 31-MAY-1996 (first entry)
XX
XX T. chalcogramma transglutaminase peptide fragment.
DE
XX transglutaminase; fish; recombinant production; Pagrus major;
KW Teragra chalcogramma; Paralichthys olivaceus; Oncorhynchus keta.
KW
XX Theragra chalcogramma.
XX
XX JP07023787-A.
PN
XX
XX 27-JAN-1995.
PD
XX
XX 13-JUL-1993; 93JP-0172998.
PF
XX
XX 13-JUL-1993; 93JP-0172998.
PR
XX
XX (AJIN) AJINOMOTO KK.
PA
XX
XX WPI; 1995-100948/14.
DR
XX
XX A fish-originated trans-glutaminase gene - can be produced in large
PT amounts at low economic cost.
PT
XX
XX Example 12; Page 42; 94pp; Japanese.
PS
XX
XX AAR65307-12 are fragments of a Theragra chalcogramma protein which has
CC

CC transglutaminase (TG) activity. The TG's of the invention can be
 CC isolated from *Pagrus major*, *Theragra chalcogramma*, *Paralichthys olivaceus*
 CC or *Oncorhynchus keta*. Fish-originated TG's are economical to produce in
 CC large quantities.
 XX
 SQ Sequence 7 AA;

Query Match 100.0%; Score 19; DB 16; Length 7;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 III
 DB 5 fnp 7

RESULT 12
 AAW52004
 ID AAW52004 standard; peptide: 7 AA.

XX AAW52004;

DT 21-SEP-1998 (first entry)

XX Peptide having immunomodulating activity.

KW Immunomodulating activity; cancer; autoimmune disorder; infection;
 KW cysteine analogue.

XX Synthetic.

FT Key Location/Qualifiers
 FT Modified-site 6
 /note= "pencillamine"

PN W09812219-A1.

XX 26-MAR-1998.

PF 15-SEP-1997; 97WO-SE01554.

PR 23-SEP-1996; 96SE-0003468.

XX (ASTR) ASTRA AB.

PI Bergstrand H, Eriksson T, Lindvall M, Saernstrand B;

DR WPI: 1998-271701/24.

PT New peptide(s) containing 3-30 amino acids - are immunomodulatory
 compounds useful in treatment of cancer or autoimmune disorders
 XX
 PS Example 138; Page 54; 75pp; English.

XX The invention relates to methionine, penicillamine and cysteine-analogue
 CC containing peptides which are immunomodulatory (either immuno-
 CC inhibitory or immunostimulatory) compounds. They may be used in treatment
 CC of cancers, infections associated with autoimmune disease, autoimmune
 CC diseases, asthma, rhinitis, fibrosis, chronic bronchitis, hepatitis,
 CC post-infectious anergy, AIDS, HIV or post-traumatic immunological
 CC anergy. The peptides may also be used as adjuvants in vaccines, or in
 CC formulations designed to prevent rejection of transplanted organs. The
 CC peptides are absorbable by the epithelial cell lining. Thus the dosage
 CC required to achieve the therapeutic effect by oral, nasal or
 CC intratracheal delivery can be much lower than that required to produce
 CC the same effect via, e.g., parenteral administration. The present
 CC sequence represents a peptide having immunomodulating activity.
 XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 19; DB 19; Length 7;

Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 III
 DB 3 fnp 5

RESULT 13
 AAW51081
 ID AAW51081 standard; peptide: 7 AA.

XX AAW51081;

DT 05-AUG-1998 (first entry)

XX Pencillamine containing peptide #44 having immuno-modulatory effect.

KW Pencillamine; immuno-modulatory; epithelial cell lining; cancer;
 KW auto-immune disease.

XX Synthetic.

FT Key Location/Qualifiers
 FT Modified-site 6
 /note= "pencillamine"

PN W09812215-A1.

XX 26-MAR-1998.

PF 15-SEP-1997; 97WO-SE01549.

PR 23-SEP-1996; 96SE-0003462.

XX (ASTR) ASTRA AB.

PI Bergstrand H, Eriksson T, Lindvall M, Saernstrand B;

DR WPI: 1998-230309/20.

PT Peptide(s) which contain penicillamine residue - are
 immunomodulatory compounds useful in treating, e.g. cancer or
 auto-immune disorders
 XX
 PS Example 42; Page 31; 71pp; English.

XX The invention relates to penicillamine containing peptides having an
 CC immuno-modulatory (either immuno-inhibitory or immuno-stimulatory)
 CC effect. The peptides may be used in treatment of cancers, infections
 CC associated with autoimmune disease, autoimmune diseases (e.g. non-obese
 CC diabetes, systemic lupus erythematosus, scleroderma, Sjogren's syndrome,
 CC dermatomyositis, multiple sclerosis, rheumatoid arthritis, arterio-
 CC sclerosis or psoriasis), asthma, rhinitis, fibrosis, chronic bronchitis,
 CC hepatitis, post-infectious anergy, AIDS, HIV or post-traumatic
 CC immunological anergy. The peptides may also be used as adjuvants in
 CC vaccines, or in formulations designed to prevent rejection of
 CC transplanted organs. The peptides are absorbable by the epithelial cell
 CC lining. Thus the dosage required to achieve the therapeutic effect by
 CC oral, nasal or intratracheal delivery can be significantly lower than
 CC that required to produce the same effect via, e.g. parenteral
 CC administration. The present sequence represents a penicillamine
 CC containing peptide.
 XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 19; DB 19; Length 7;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 III

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:10:39 ; Search time 100.84 Seconds
(without alignments)
2.859 Million cell updates/sec

Title: US-09-824-286-17

Perfect score: 19

Sequence: 1 FNP 3

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	15	2 S08209	hypothetical prote
2	19	100.0	15	2 A61612	allatostatin - tob
3	19	100.0	20	2 A54519	tubulin alpha chain
4	19	100.0	24	2 T45851	adenosylmethionine
5	19	100.0	25	2 T46841	hypothetical prote
6	19	100.0	25	2 S69139	bumetanide-binding
7	19	100.0	26	2 PS0111	beta-galactoside b
8	19	100.0	27	2 H44616	homeotic protein H
9	19	100.0	27	2 T44616	homeotic protein H
10	19	100.0	29	2 A55445	repy protein - Esc
11	19	100.0	31	2 G82278	hypothetical prote
12	19	100.0	33	2 S65599	hevein - Para rubb
13	19	100.0	35	2 S42876	probable succinate
14	19	100.0	35	2 H82103	hypothetical prote
15	19	100.0	35	2 A55850	7alpha-hydroxyster
16	19	100.0	37	2 B45187	homeotic protein G
17	19	100.0	37	2 D95008	hypothetical prote
18	19	100.0	40	2 S77793	transketolase (EC
19	19	100.0	40	2 B24095	gamma-hordein 3 -
20	19	100.0	40	2 T06840	photosystem I chain
21	19	100.0	42	2 A05049	hypothetical prote
22	19	100.0	45	2 C82233	hypothetical prote
23	19	100.0	45	2 E49050	T-cell surface gly
24	19	100.0	46	2 S12084	hypothetical prote
25	19	100.0	48	2 P00617	DNA-directed DNA p
26	19	100.0	49	2 S69087	annexin VI, mitoch
27	19	100.0	49	2 AG1063	hypothetical secre
28	19	100.0	50	2 F90761	hypothetical prote
29	19	100.0	51	1 A61125	insulin - American

30	19	100.0	51	1 R6UC46	ribosomal protein
31	19	100.0	51	2 C46353	ORF3 protein - coc
32	19	100.0	52	2 S06492	beta-galactoside-b
33	19	100.0	52	2 S63981	immune-inducible p
34	19	100.0	53	2 S66603	xanthine oxidase (
35	19	100.0	53	2 S53698	ribosomal protein
36	19	100.0	53	2 B55392	abacsin precursor
37	19	100.0	53	2 A96216	hypothetical prote
38	19	100.0	54	2 A72206	hypothetical prote
39	19	100.0	55	2 E90626	ATP synthase F0 ch
40	19	100.0	55	2 D37334	D-mandelate dehydr
41	19	100.0	55	2 AD2382	hypothetical prote
42	19	100.0	55	2 G98149	hypothetical prote
43	19	100.0	56	2 P50424	H11 protein - Toxo
44	19	100.0	57	2 T03092	hypothetical prote
45	19	100.0	57	2 T22546	hypothetical prote

ALIGNMENTS

RESULT 1
S08209
hypothetical protein 2 - garden pea
N:Alternate names: phytochrome
C:Species: Pisum sativum (garden pea)
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Sep-1997
C:Accession: S08209
R:Sato, N.
Plant Mol. Biol. 11, 697-710, 1988
A:Title: Nucleotide sequence and expression of the phytochrome gene in Pisum sativum
A:Reference number: S06856
A:Accession: S08209
A:Molecule type: DNA
A:Residues: 1-15 <SAT>
A:Cross-references: EMBL:X14077; NID:q20836; PID:q20838
A:Genetics:
A:Gene: phy

Query Match 100.0%; Score 19; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 FNP 3
DB 11 FNP 13
RESULT 2
A61612
allatostatin - tobacco hornworm
C:Species: Manduca sexta (tobacco hornworm)
C>Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 11-Jul-1997
C:Accession: A61612
R:Kramer, S.J.; Toschi, A.; Miller, C.A.; Kataoka, H.; Quistad, G.B.; Li, J.P.; Carn
Proc. Natl. Acad. Sci. U.S.A. 88, 9458-9462, 1991
A:Title: Identification of an allatostatin from the tobacco hornworm Manduca sexta.
A:Reference number: A61612; MUID:92052112
A:Accession: A61612
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-15 <KRA>
C:Keywords: neuropeptide; pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 100.0%; Score 19; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 FNP 3
DB 11

Db 9 FNP 11

RESULT 3
A54519
tubulin alpha chain - Leishmania enriettii (fragment)
C:Species: Leishmania enriettii
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 13-Aug-1999
C:Accession: A54519
R:Wirth, D.F.; Slater, C.
Mol. Biochem. Parasitol. 9, 83-92, 1983
A:Title: Isolation and characterization of an alpha-tubulin gene from Leishmania enriettii
A:Reference number: A54519; MUID:84142075
A:Accession: A54519
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-20 <MIR>
A:Cross-references: GB:M28001; NID:g159409; PIDN:AAA29273.1; PID:g159410
C:Superfamily: tubulin

Query Match 100.0%; Score 19; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 FNP 3
|||
Db 17 FNP 19

RESULT 4
I45851
adenosylmethionine decarboxylase (EC 4.1.1.50) precursor - bovine (fragment)
N:Alternate names: Adomet decarboxylase
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 22-Jun-1999
C:Accession: I45851
R:Mach, M.; White, M.W.; Neubauer, M.G.; Degen, J.L.; Morris, D.R.
J. Biol. Chem. 261, 11697-11703, 1986
A:Title: Isolation of a cDNA clone encoding S-adenosylmethionine decarboxylase: expression
A:Reference number: I45851; MUID:86304300
A:Accession: I45851
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-24 <MAC>
A:Cross-references: GB:M14289; NID:g162624; PIDN:AAA30360.1; PID:g552313
C:Superfamily: eukaryotic adenosylmethionine decarboxylase
C:Keywords: carbon-carbon lyase; carboxy-lyase; polyamine biosynthesis

Query Match 100.0%; Score 19; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
|||
Db 15 FNP 17

RESULT 5
T46841
hypothetical protein [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C:Accession: T46841
R:Treuner-Lange, A.; Kuhn, A.; Duerre, P.
J. Bacteriol. 179, 4501-4512, 1997
A:Title: The kdp system of Clostridium acetobutylicum: cloning, sequencing, and transcription
A:Reference number: Z24106; MUID:97369818
A:Accession: T46841
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-25 <TRD>

A:Cross-references: EMBL:U44892; NID:g2275246; PIDN:AAC45476.1; PID:g2275248
A:Experimental source: DSM 79

Query Match 100.0%; Score 19; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
|||
Db 20 FNP 22

RESULT 6
S69139
bumetanide-binding protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 17-Mar-1999
C:Accession: S69139
R:Otallah-Kolac, M.; Tripier, D.; Bruehl, B.; Platte, H.D.; Jouvenal, K.; Schuh, K.;
Eur. J. Biochem. 228, 506-514, 1995
A:Title: The 60-kDa bumetanide-binding protein from rat liver membranes is a catalase
A:Reference number: S69139; MUID:95220382
A:Accession: S69139
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-25 <OTY>

Query Match 100.0%; Score 19; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
|||
Db 19 FNP 21

RESULT 7
PS0111
beta-galactoside binding lectin - common marmoset (fragments)
C:Species: Callithrix jacchus (common marmoset)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 07-May-1999
C:Accession: PS0111
R:Ohnawa, F.; Hirano, F.; Natori, S.
J. Biochem. 107, 431-434, 1990
A:Title: Purification and properties of a beta-galactoside-binding lectin from neonat
A:Reference number: PS0111; MUID:90256719
A:Accession: PS0111
A:Molecule type: Protein
A:Residues: 1-26 <OHS>

Query Match 100.0%; Score 19; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
|||
Db 24 FNP 26

RESULT 8
H44616
hemectic protein Hox 8 (clone 8q) - sea lamprey (fragment)
C:Species: Petromyzon marinus (sea lamprey)
C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 15-Oct-1999
C:Accession: H44616
R:Pendleton, J.W.; Nagai, B.K.; Murtha, M.T.; Ruddle, F.H.
Proc. Natl. Acad. Sci. U.S.A. 90, 6300-6304, 1993
A:Title: Expansion of the Hox gene family and the evolution of chordates.
A:Reference number: A44616; MUID:93317669
A:Accession: H44616

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-27 <PEN>
A:Cross-references: GB:U14901; NID:g290821; PIDN:AAA02536.1; PID:g290822
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation

Query Match 100.0%; Score 19; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
|||
Db 2 FNP 4

RESULT 9
144616
homeotic protein Hox 8 (clone 8r) - sea lamprey (fragment)
C:Species: Petromyzon marinus (sea lamprey)
C:Date: 07-Oct-1994 #sequence-revision 07-Oct-1994 #text-change 15-Oct-1999
C:Accession: U144616
R:Pendleton, J.W.; Nagai, B.K.; Murtha, M.T.; Ruddle, F.H.
Proc. Natl. Acad. Sci. U.S.A. 90, 6300-6304, 1993
A:Title: Expansion of the Hox gene family and the evolution of chordates.
A:Reference number: A44616; MUID:93317669
A:Accession: U144616
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-27 <PEN>
A:Cross-references: GB:U14903; NID:g290823; PIDN:AAA02538.1; PID:g290824
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation

Query Match 100.0%; Score 19; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
|||
Db 2 FNP 4

RESULT 10
A35445
repl protein - Escherichia coli plasmids
N:Alternate names: repB protein
C:Species: Escherichia coli
C:Date: 14-Sep-1990 #sequence-revision 14-Sep-1990 #text-change 19-Oct-1995
C:Accession: A35445; MUID:90285201
R:Hama, C.; Takizawa, T.; Moriwaki, H.; Mizobuchi, K.
J. Biol. Chem. 265, 10666-10673, 1990
A:Title: Role of leader peptide synthesis in repZ gene expression of the Colib-P9 plasmid.
A:Reference number: A35445; MUID:90285201
A:Accession: A35445
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-29 <HAM>
A:Experimental source: plasmid Colib-P9
J. Paszkier, J.; Wilson, I.W.; Piltard, A.J.
J. Bacteriol. 174, 2376-2383, 1992
A:Title: Mutations affecting translational coupling between the rep genes of an IncB min
A:Reference number: A42382; MUID:92202169
A:Accession: A42382
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-29 <PRA>
A:Cross-references: GB:M93063; GB:M38522
A:Experimental source: plasmid pM0720
C:Genetics:
A:Genome: plasmid

Query Match 100.0%; Score 19; DB 2; Length 29;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
|||
Db 7 FNP 9

RESULT 11
G82278
hypothetical protein VC0805 [Imported] - Vibrio cholerae (strain N16961 serogroup O1
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence-revision 20-Aug-2000 #text-change 02-Feb-2001
C:Accession: G82278
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: G82278
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-31 <HEI>
A:Cross-references: GB:AE004166; GB:AE003852; NID:g96555259; PIDN:AAF93969.1; GSPDB:G
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0805
A:Map position: 1

Query Match 100.0%; Score 19; DB 2; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
|||
Db 17 FNP 19

RESULT 12
S65599
hevein - Para rubber tree (fragments)
C:Species: Hevea brasiliensis (Para rubber tree)
C:Date: 28-Oct-1996 #sequence-revision 13-Mar-1997 #text-change 07-May-1999
C:Accession: S65599
R:Soedjanaatmadja, U.M.S.; Subroto, T.; Beintema, J.J.
FEBS Lett. 363, 211-213, 1995
A:Title: Processed products of the hevein precursor in the latex of the rubber tree (C
A:Reference number: S65599; MUID:95253547
A:Accession: S65599
A:Status: preliminary
A:Molecule type: Protein
A:Residues: 1-15; 16-33 <SOE>
C:Superfamily: hevein precursor; barwin homology; hevein chitin-binding domain homol

Query Match 100.0%; Score 19; DB 2; Length 33;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
|||
Db 20 FNP 22

RESULT 13
S42876
probable succinate--CoA lyase (ADP-forming) (EC 6.2.1.5) beta chain - Coxiella burn
N:Alternate names: probable succinyl-CoA synthetase beta chain

C:Species: *Coxiella burnetii*
 C:Date: 20-Feb-1995 #sequence_revision 26-Jul-1996 #text_change 05-May-2000
 C:Accession: S42876
 R:Title: D.; Willems, H.; Oswald, W.; Krauss, H.
 submitted to the EMBL Data Library, March 1994
 A:Reference number: S42872
 A:Accession: S42876
 A:Molecule type: DNA
 A:Residues: 1-35 <THI>
 A:Cross-references: EMBL:X77919
 C:Superfamily: succinate--CoA ligase (ADP-forming) beta chain
 C:Keywords: acid-thiol ligase; coenzyme A

Query Match 100.0%; Score 19; DB 2; Length 35;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 FNP 3
 Db 26 FNP 28

RESULT 14
 H82103
 hypothetical protein VC2228 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)
 C:Species: *Vibrio cholerae*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: H82103
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
 chardson, D.; Ermolaeva, M.D.; Veluthavan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, F.
 l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406:477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A:Reference number: A82035; MUID:20406833
 A:Accession: H82103
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-35 <HEI>
 A:Cross-references: GB:AE04294; GB:AE03852; NID:96656774; PIDN:AAF95372.1; GSPDB:GN001
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC2228
 A:Map position: 1

Query Match 100.0%; Score 19; DB 2; Length 35;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FNP 3
 Db 4 FNP 6

RESULT 15
 A55850
 7alpha-hydroxysteroid dehydrogenase (EC 1.1.1.159) - *Clostridium sordellii*
 C:Species: *Clostridium sordellii*
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 15-Oct-1999
 C:Accession: A55850
 R:Coleman, J.P.; Hudson, L.L.; Adams, M.J.
 J. Bacteriol. 176: 4865-4874, 1994
 A:Title: Characterization and regulation of the NADP-linked 7alpha-hydroxysteroid dehydrogenase
 A:Reference number: A55850; MUID:94327451
 A:Accession: A55850
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-35 <COL>
 A:Cross-references: GB:L12058; NID:9431280; PIDN:AAA5355.1; PID:9577669
 C:Keywords: oxidoreductase

Query Match 100.0%; Score 19; DB 2; Length 35;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 FNP 3
 Db 12 FNP 14

Search completed: August 20, 2002, 11:10:41
 Job time: 5671 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:33:03 ; Search time 55.29 Seconds

(Without alignments)
2.101 Million cell updates/sec

Title: US-09-824-286-17

Perfect score: 19

Sequence: 1 FNP 3

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	15	ALIS_MANSE	P42559 manduca sex
2	19	100.0	15	RM12_YEAST	P36522 saccharomyc
3	19	100.0	39	ABAE_BOMPA	P81463 bombus pasc
4	19	100.0	40	PSAI_CYAPA	P48117 cyanophora
5	19	100.0	46	COAT_BPH75	P82889 bacterioph
6	19	100.0	46	YPCA_ECOLI	P19755 escherichia
7	19	100.0	50	RL39_SUDAC	P13005 sulfolobus
8	19	100.0	51	INS_ANGRO	P42633 anguilla ro
9	19	100.0	52	MTK_DROME	O24395 drosophila
10	19	100.0	53	ABAE_APIME	P15450 apis mellif
11	19	100.0	55	ITRP_HALRO	P16589 halocynthia
12	19	100.0	61	PSBK_HORVU	P25877 hordeum vul
13	19	100.0	61	PSBK_MAIZE	P48188 zea mays (m
14	19	100.0	61	PSBK_ORYSA	P12162 oryza sativ
15	19	100.0	63	PSBK_WHEAT	P58273 triticum ae
16	19	100.0	63	RPCX_SCHPO	P48011 schizosacch
17	19	100.0	64	V592_HAEIN	P44021 haemophilus
18	19	100.0	66	ATP8_CEREH	O2mq19 cervus elap
19	19	100.0	68	YE02_HAEIN	P44177 haemophilus
20	19	100.0	70	CSP1_ECOLI	P77605 salmonella
21	19	100.0	70	CSP1_SALTI	P39818 salmonella
22	19	100.0	74	HXBR_CHICK	P23681 gallus gall
23	19	100.0	74	NUML_MOUSE	O62425 mus musculu
24	19	100.0	77	ICIC_HORVU	P01054 hordeum vul
25	19	100.0	79	RS16_BUCAI	P57474 buchnera ap
26	19	100.0	81	IAC_MACFA	P34953 macaca fasc
27	19	100.0	81	NUML_HUMAN	O00483 homo sapien
28	19	100.0	82	NUML_BOVIN	O01331 bos taurus
29	19	100.0	82	RS16_ECOLI	P02372 escherichia
30	19	100.0	82	RS16_HAEIN	P44382 haemophilus
31	19	100.0	82	RS16_PASMU	P58123 pasteurella
32	19	100.0	82	RS16_PASMU	P58123 pasteurella
33	19	100.0	82	RS16_SALTY	P36242 salmonella

34	19	100.0	82	1	RS16_VIBCH	O9kug0 vibrio chol
35	19	100.0	82	1	SIRA_VIBCH	O9kvw4 vibrio chol
36	19	100.0	83	1	ICIB_HORVU	P10663 hordeum vul
37	19	100.0	83	1	RS16_PSBAB	O9hxp9 pseudomonas
38	19	100.0	84	1	IAC2_HUMAN	P20155 homo sapien
39	19	100.0	86	1	ILVM_ECOLI	P13048 escherichia
40	19	100.0	86	1	YIDD_PROMI	P22834 proteus mir
41	19	100.0	88	1	VG30_BPM2	O64224 mycobacteri
42	19	100.0	88	1	VYAD_VACC	P20513 vaccinia vl
43	19	100.0	92	1	HXB8_RAT	P18863 rattus norv
44	19	100.0	92	1	V766_RICPR	O9zch6 rickettsia
45	19	100.0	94	1	F16B_MOUSE	P97323 mus musculu

ALIGNMENTS

RESULT	1	ALIS_MANSE	STANDARD	PRT	15 AA.
ID	ALIS_MANSE				
AC	P42559				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	Allatostatin (Mas-AS).				
OS	Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;				
OC	Sphingioidea; Sphingidae; Sphinginae; Manduca.				
OX	NCBI_TaxID=7130;				
RN	[1]				
RP	SEQUENCE.				
RC	TISSUE=Head;				
RX	MEDLINE=92052112; PubMed=1946359;				
RA	Kramer S.-J., Toschl A., Miller C.A., Kataoka H., Quistad G.B.,				
RA	L.J.P., Carney R.L., Schooley D.A.;				
RT	"Identification of an allatostatin from the tobacco hornworm Manduca				
RT	sexta.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 88:9458-9462(1991).				
CC	-1- FUNCTION: STRONGLY INHIBITS JUVENILE HORMONE BIOSYNTHESIS IN VITRO				
CC	BY THE CORPORA ALLATA FROM FIFTH-STADIUM LARVAE AND ADULT FEMALES.				
CC	-1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.				
FT	Neuropeptide.				
FT	MOD_RES				
FT	SEQUENCE				
SO	SEQUENCE				
QY	1 FNP 3				
Db	9 FNP 11				
Query Match	100.0%;				
Best Local Similarity	100.0%;				
Matches	3; Conservative				
Mismatches	0; Indels				
Gaps	0;				
RESULT	2				
ID	RM12_YEAST				
AC	P36522				
DT	01-JUN-1994 (Rel. 29, Created)				
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DT	01-JUN-1994 (Rel. 29, Last annotation update)				
DE	Mitochondrial 60S ribosomal protein L12 (Iml12) (Fragment).				
GN	MRP12.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE.				
RX	MEDLINE=91285106; PubMed=2060626;				
RA	Grohmann L., Graack H.-R., Kruff V., Choll T., Goldschmidt-Reisin S.,				

```

RA Kitakawa M.;
RT "extended N-terminal sequencing of proteins of the large ribosomal
RL subunit from yeast mitochondria.";
DR FEBS Lett. 284:51-56(1991).
RL PIR: S17261; S17261.
DR SGD: L0002667; MRPL12.
KM Ribosomal protein; Mitochondrion.
FT NON_TER
SQ SEQUENCE 15 AA: 1851 MW; 74BCD9FEDDB3900 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
DB 13 FNP 15

RESULT 3
ABAE_BOMPA STANDARD; PRT; 39 AA.
AC P81463;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE ABAecin.
OS Bombus pascuorum.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Prexygota; Neoptera; Endopterygota; Hymenoptera; Apoecrita; Aculeata;
OC Apoidea; Apidae; Bombs.
ON NCBI_Taxid=65598;
RN [1]
RP SEQUENCE.
RC TISSUE=Hemolymph;
RX MEDLINE=97362903; Pubmed=9219367;
RA Rees J.A., Montatte M., Bulet P.;
RT "Novel antibacterial peptides isolated from a European bumblebee,
RL Bombus pascuorum (Hymenoptera, Apoidea).";
RL Insect Biochem. Mol. Biol. 27:413-422(1997).
CC -1- FUNCTION: ANTIBACTERIAL PEPTIDE AGAINST GRAM-POSITIVE AND
CC GRAM-NEGATIVE BACTERIA.
CC -1- INDUCTION: BY BACTERIAL INFECTION.
CC -1- SIMILARITY: PARTIAL TO APIDAEICINS AND DIPTERICINS.
KM Insect immunity; Antibiotic; Hemolymph.
SQ SEQUENCE 39 AA: 4395 MW; 52E952E25D13A097 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
DB 24 FNP 26

RESULT 4
PSAJ_CYAPA STANDARD; PRT; 40 AA.
ID PSAJ_CYAPA
AC P84117;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Photosystem I reaction center subunit IX (PSI-J).
GN PSAJ.
OS Cyanophora paradoxa.
OG Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_Taxid=2762;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=LB555 / PRINGSHEIM;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohmert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LB555 / PRINGSHEIM;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohmert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -1- FUNCTION: MAY HELP IN THE ORGANIZATION OF THE PSAE AND PSAF
CC SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE PSAJ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U00821; AAA81183.1; -
DR Mendel: 7884; CYAPA:psaj.1.
DR InterPro: IPR002615; PSI_PSAJ.
DR Pfam: PF01701; PSI_PSAJ; 1.
DR ProDom: PD004198; PSI_PSAJ; 1.
DR Cyanelle: Photosystem I; Photosynthesis; Transmembrane.
FT TRANSMEM
SQ SEQUENCE 40 AA: 4481 MW; 798A8CA36A007153 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
DB 31 FNP 33

RESULT 5
COAT_BPH75 STANDARD; PRT; 46 AA.
ID COAT_BPH75
AC P82889;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein P8 (Major coat protein).
OS Bacteriophage PH75.
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
OX NCBI_Taxid=144736;
RN [1]
RP SEQUENCE, AND FORMYLATION.
RX MEDLINE=21265138; Pubmed=11371161;
RA Pederson D.M., Welsh L.C., Marvin D.A., Sampson M., Perham R.N.,
RA Yu M., Slater M.R.;
RT "The protein capsid of filamentous bacteriophage PH75 from Thermus
RT thermophilus.";
RL J. Mol. Biol. 309:401-421(2001).
CC -1- FUNCTION: COAT PROTEIN P8 IS THE MAJOR COAT PROTEIN OF THE
CC VIRION.
CC -1- SUBUNIT: THERE ARE SEVERAL COPIES OF THIS PROTEIN IN THE COAT OF
CC PHAGE PH75. THEY ARE ARRANGED IN A HELIX AROUND THE DNA.
CC -1- SUBCELLULAR LOCATION: Type III membrane protein (Potential).
KM Coat protein; Formylation; Transmembrane.

```

FT DOMAIN 1 15 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 16 36 POTENTIAL.
FT DOMAIN 37 46 INTRACELLULAR (POTENTIAL).
FT MOD_RES 1 1 FORMYLATION.
SQ SEQUENCE 46 AA; 4813 MW; 1E2EBFA950C7ACBE CRC64;

Query Match 100.0%; Score 19; DB 1; Length 46;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
Db 3 FNP 5

RESULT 6
YPC4_ECOLI STANDARD; PRT; 46 AA.

AC P19755;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1991 (Rel. 20, Last annotation update)
DE Hypothetical 5.3 kDa protein (ORF 46).
OS Escherichia coli.
OC Plasmid Incn pcu1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_Taxid=562;

RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=K12;

RX MEDLINE=90382682; PubMed=2205534;
RA Krishnan B.R., Robert P.R., Seltzer U., Iyer V.N.;
RT "Mutations within the replicon of the Incn plasmid pcu1 that affect
its Escherichia coli polA-independence but not its autonomous
RT replication ability.";
RL Gene 91:1-7(1990).

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M18262; AAA98067.1; -.
DR PIR: S12084; S12084.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 46 AA; 5304 MW; 14F81E5F636952B2 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 46;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
Db 35 FNP 37

RESULT 7
RL39_SULIAC STANDARD; PRT; 50 AA.

AC P13005; P95686;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L39e (L46e).
GN RPL39E OR RPL46 OR RIL46.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.

OX NCBI_Taxid=2285;
RN 11
RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-46.

RC STRAIN=ATCC 35091 / DSM 1616 / PL;
RX MEDLINE=89325607; PubMed=2502431;
RA Ramirez C., Louie K.A., Matheson A.T.;
RT "A small basic ribosomal protein in Sulfolobus solfataricus
RT equivalent to L46 in yeast: structure of the protein and its gene.";
RL FEBS Lett. 250:416-418(1989).

RN 12
RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=95226466; PubMed=7711082;
RA Moll R., Schmidtke S., Schaefer G.;
RT "Nucleotide sequence of a gene cluster encoding ribosomal proteins in
RT the thermophilic crenarchaeon Sulfolobus acidocaldarius.";
RL Biochim. Biophys. Acta 1261:315-318(1995).
CC -1- FUNCTION: BINDS SPECIFICALLY TO A REGION IN 26S RNA.
CC NEAR THE SUBUNIT INTERFACE.

CC -1- SIMILARITY: BELONGS TO THE L39E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X16161; CAA34287.1; -.
DR EMBL: X77509; CAA54638.1; ALT_INIT.

DR PIR: S05009; R6UC46.
DR PIR: S41963; S41963.

DR InterPro: IPR000077; Ribosomal_L39.
DR Pfam: PF00832; Ribosomal_L39; 1.
DR ProDom: PD007914; Ribosomal_L39; FALSE_NEG.
DR PROSITE: PS00051; RIBOSOMAL_L39E; 1.
KW Ribosomal protein.
FT INIT MET 0
SQ SEQUENCE 50 AA; 5932 MW; 217CB51B357B44F9 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
Db 36 FNP 38

RESULT 8
INS_ANGRO STANDARD; PRT; 51 AA.

AC P42633;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin.
GN INS.
OS Anguilla rostrata (American eel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;
OC Anguillidae; Anguilla.
OX NCBI_Taxid=7938;

RN 11
RP SEQUENCE.

RC TISSUE=Pancreas;
RX MEDLINE=91340068; PubMed=1874385;
RA Conlon J.M., Andrews P.C., Thim L., Moon T.W.;
RT "The primary structure of glucagon-like peptide but not insulin has
RT been conserved between the American eel, Anguilla rostrata and the
RT European eel, Anguilla anguilla.";

FL Gen. Comp. Endocrinol. 82:23-32(1991).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HEMERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR HSSP: P01308; 1HIS.
DR InterPro: IPR000739; Insulin_IGF_Relaxin.
DR Pfam: PF000649; Insulin_1.
DR PRINTS: PR00276; INSULIN.
DR PRINTS: PR00277; INSULIN.
DR SMART: SM00078; 1IGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5652 MW; 1999FD7EBA173CB2 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
Db 26 FNP 28

RESULT 9

MTK_DROME STANDARD; PRT: 52 AA.
AC Q24395; Q24396; Q9Y7B9;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Melchikowin precursor.
GN MTK OR CG8175.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 27-52.
RC STRAIN=OREGON-R; TISSUE=Abdomen, and Thorax;
RX MEDLINE=96067716; PubMed=758819;
RA Levashtina E.A., Ohresser S., Bulat B., Reichart J.-M., Hetru C.,
RA Hoffmann J.A.;
RT "Metchnikowin, a novel immune-inducible proline-rich peptide from
RT Drosophila with antibacterial and antifungal properties.";
RL Eur. J. Biochem. 233:694-700(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=98263241; PubMed=9600835;
RA Levashtina E.A., Ohresser S., Lemaitre B., Imler J.-L.;
RT "Two distinct pathways can control expression of the gene encoding
RT the Drosophila antimicrobial peptide metchnikowin.";
RL J. Mol. Biol. 278:515-527(1998).
RN [3]

RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abdill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler A.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Daventport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liao P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q., Zheng L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT Science 287:2185-2195(2000).
CC -1- FUNCTION: POTENT ANTIFUNGAL AND ANTIBACTERIAL ACTIVITY AGAINST
CC GRAM-POSITIVE BACTERIA.
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN FAT BODY. MAYBE BLOOD
CC CELLS PARTICIPATE IN THE PRODUCTION OF THE PEPTIDE.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED RAPIDLY AND STRONGLY AT ALL STAGES.
CC -1- POLYMORPHISM: 2 ALLELIC FORMS (A1 AND A2) VARYING IN TWO AA
CC POSITION. THE ISOFORM SHOWN HERE IS A1.
CC -1- SIMILARITY: TO DIPERICIN, HEMIPERICIN, DROSOCIN, APIDECINS AND
CC TO THE C-TERMINUS OF ABACIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X91060; CAA62511.1; -;
DR EMBL: X91061; CAA62512.1; -;
DR EMBL: AF030959; AAC64659.1; -;
DR EMBL: AE003811; AAF58139.1; -;
DR FLYBase: FBgn0014865; Mtk.
KW Insect immunity; Antibiotic; Fungicide; Signal.
FT SIGNAL 1 24
FT PROPEP 25 26
FT PEPTIDE 27 52 METCHNIKOWIN.
FT VARIANT 19 19 A -> T (IN STRAIN A2).
FT VARIANT 29 29 H -> R (IN STRAIN A2).
SQ SEQUENCE 52 AA; 5654 MW; 0B1E2112BAR03129 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
Db 26 FNP 28

Db 41 FNP 43

RESULT 10
ID ABAE_APIME STANDARD; PRT; 53 AA.
AC P15450;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Abaecin precursor.
OS Apis mellifera (Honeybee).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Apoidea; Apidae; Apis.
NCBI_TaxID=7460;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95050655; Pubmed=7961803;
RA Casteels-Josson K., Zhang W., Capaci T., Tempst P.;
RT "Acute transcriptional response of the honeybee peptide-antibiotics
RT gene repertoire and required post-translational conversion of the
RT precursor structures.";
RL J. Biol. Chem. 269:28569-28575(1994).
RN [2]
RP SEQUENCE OF 20-53.
RC TISSUE=Hemolymph;
RX MEDLINE=90126848; Pubmed=2298215;
RA Casteels P., Ampe C., Riviere L., van Damme J., Elicone C., Jacobs F.,
RT Tempst P.;
RT "Isolation and characterization of abaecin, a major antibacterial
RT response peptide in the honeybee (*Apis mellifera*).";
RL Eur. J. Biochem. 187:381-386(1990).
CC -1- FUNCTION: THIS PEPTIDE HAS BACTERICIDAL ACTIVITY.
CC -1- SIMILARITY: PARTIAL TO APIDACINS AND DIPERICINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U15954; AAA67442.1; -
DR PIR: S08152; S08152.
KW Insect immunity; Antibiotic; Hemolymph; Signal.
FT SIGNAL 1 19
FT CHAIN 20 53 ABAECIN.
FT VARIANT 52 52 G->S.
SO SEQUENCE 53 AA; 5903 MW; 0EB1F0CC57797FBC CRC64;

Query Match 100.0%; Score 19; DB 1; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FNP 3
Db 43 FNP 45

RESULT 11
ID ITRP_HALRO STANDARD; PRT; 55 AA.
AC P16589;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trypsin inhibitor.
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Pyrosidae; Halocynthia.

OX NCBI_TaxID=7729;
RN [1]
RP SEQUENCE.
RC TISSUE=Hemolymph;
RX MEDLINE=90256715; Pubmed=2341375;
RA Kunazaki T., Hoshida N., Yokosawa H., Ishii S.-I.;
RT "Primary structure of ascidian trypsin inhibitors in the hemolymph of
RT a solitary ascidian, *Halocynthia roretzi*.";
RL J. Biochem. 107:409-413(1990).
RN [2]
RP DISULFIDE BONDS.
RX MEDLINE=90256716; Pubmed=2111316;
RA Kunazaki T., Ishii S.-I.;
RT "Disulfide bridge structure of ascidian trypsin inhibitor I:
RT similarity to Kazal-type inhibitors.";
RL J. Biochem. 107:414-419(1990).
CC -1- FUNCTION: POTENT INHIBITOR OF TRYPSIN.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: SOME TOPOLOGICAL SIMILARITY TO KAZAL-TYPE INHIBITORS.
KW Serine protease inhibitor; Hemolymph.
FT DISULFID 5 40
FT DISULFID 12 41
FT DISULFID 14 37
FT DISULFID 23 34
FT ACT_SITE 16 17 REACTIVE BOND.
SO SEQUENCE 55 AA; 6078 MW; 412154BDB33207C2 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 55;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FNP 3
Db 8 FNP 10

RESULT 12
ID PSBK_HORVU STANDARD; PRT; 61 AA.
AC P25877;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II reaction center protein K precursor (PSII-K).
GN PSBK.
OS Hordeum vulgare (Barley), and
OS Secale cereale (Rye).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Hordeum.
NCBI_TaxID=4513, 4550;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=H. vulgare; TISSUE=Seedling;
RX MEDLINE=90291518; Pubmed=1694111;
RA Sexton T.B., Jones J.T., Mulliet J.E.;
RT "Sequence and transcriptional analysis of the barley ctDNA region
RT upstream of psbD-psbC encoding trnK(uru), trp54, trnD(uuc), psbK,
RT psbI, and trnS(gcu).";
RL Curr. Genet. 17:445-454(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S. cereale; Pubmed=2644624;
RX MEDLINE=89128450;
RA Burkhart A.A., Kolosov V.L., Klevovich O.N., Zolotarev A.S.;
RT "Nucleotide sequence of rye chloroplast DNA fragment, comprising
RT psbD, psbC and trnS genes.";
RL Nucleic Acids Res. 17:798-798(1989).
CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER
CC OF PHOTOSYSTEM II.
CC -1- SIMILARITY: BELONGS TO THE PSBK FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X52765; CAA36975.1; -
DR EMBL: X61674; CAA43848.1; -
DR PIR: S28768; S28768.
DR PIR: S31447; S31447.
DR InterPro: IPR003687; PsbK.
DR Pfam: PF02533; PsbK; 1.
KW Photosystem II; Chloroplast.
FT PROPEP 1 24
FT CHAIN 25 61
SQ SEQUENCE 61 AA; 6996 MW; 84C901067C454777 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 61;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
DB 33 FNP 35

RESULT 13
PSBK_MAIZE STANDARD; PRT; 61 AA.
ID PSBK_MAIZE
AC P48188;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II reaction center protein K precursor (PsbK).
GN Zea mays (Maize).
OS Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95395841; PubMed=7666415;
RA Maier R.M., Neckermann K., Igloi G.L., Koessel H.;
RT "Complete sequence of the maize chloroplast genome: gene content,
RT hotspots of divergence and fine tuning of genetic information by
RT transcript editing."
RL J. Mol. Biol. 251:614-628(1995).
CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER
CC OF PHOTOSYSTEM II.
CC -1- SIMILARITY: BELONGS TO THE PSBK FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X86563; CAA60268.1; -
DR MaizeDB: 118217; -
DR Mendel: 2686; ZEma;psbk;1.
DR InterPro: IPR003687; PsbK.
DR Pfam: PF02533; PsbK; 1.
KW Photosystem II; Chloroplast.
FT PROPEP 1 24
FT CHAIN 25 61
SQ SEQUENCE 61 AA; 6982 MW; BHC361067C4158F5 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 61;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
DB 33 FNP 35

RESULT 15
PSBK_WHEAT STANDARD; PRT; 61 AA.
ID PSBK_WHEAT

SQ SEQUENCE 61 AA; 6874 MW; D67361066179A2C1 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 61;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
DB 33 FNP 35

RESULT 14
PSBK_ORYZA STANDARD; PRT; 61 AA.
ID PSBK_ORYZA
AC P12162;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II reaction center protein K precursor (PsbK).
GN Oryza sativa (Rice).
OS Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RX MEDLINE=89364698; PubMed=2770692;
RA Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M.,
RA Mori M., Kondo C., Honji Y., Sun C.-R., Meng B.-Y., Li Y.-Q.,
RA Kanno A., Nishizawa Y., Hirai A., Shinozaki K., Sugita M.;
RT "The complete sequence of the rice (Oryza sativa) chloroplast genome:
RT intermolecular recombination between distinct trna genes accounts for
RT a major plastid DNA inversion during the evolution of the cereals."
RL Mol. Gen. Genet. 217:185-194(1989).
CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER
CC OF PHOTOSYSTEM II.
CC -1- SIMILARITY: BELONGS TO THE PSBK FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X15901; CAA34010.1; -
DR PIR: J00203; F2R2KS.
DR Mendel: 2688; ORYsa;psbk;1.
DR InterPro: IPR003687; PsbK.
DR Pfam: PF02533; PsbK; 1.
KW Photosystem II; Chloroplast.
FT PROPEP 1 24
FT CHAIN 25 61
SQ SEQUENCE 61 AA; 6982 MW; BHC361067C4158F5 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 61;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
DB 33 FNP 35

RESULT 15
PSBK_WHEAT STANDARD; PRT; 61 AA.
ID PSBK_WHEAT

AC P58273;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Photosystem II reaction center protein K precursor (PSII-K).
 GN PSBK.
 OS *Triticum aestivum* (Wheat).
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; *Triticum*.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPRAIN-cv. Chinese Spring;
 RA Ogihara Y., Isono K., Kojima T., Endo A., Hanaoka M., Shina T.,
 RA Terachi T., Utsugi S., Murata M., Mori N., Takumi S., Ikeo K.,
 RA Gojobori T., Murai R., Murai K., Matsuoka Y., Ohnishi Y., Tajiri H.,
 RA Tsunewaki K.;
 RT "Chinese spring wheat (*Triticum aestivum* L.) chloroplast genome:
 RT complete sequence and contig clones."
 RL Plant Mol. Biol. Rep. 18:243-253(2000)
 CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER
 CC OF PHOTOSYSTEM II.
 CC -!- SIMILARITY: BELONGS TO THE PSBK FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB042240; BAH47017.1; -;
 DR InterPro: IPR003687; PSBK.
 DR Pfam: PF02533; PSBK; 1.
 KW Photosystem II; Chloroplast.
 FT PROPEP 1 24 BY SIMILARITY.
 FT CHAIN 25 61 PHOTOSYSTEM II REACTION CENTER PROTEIN K.
 SO SEQUENCE 61 AA; 7030 MW; 467901067C4158FD CRC64;

Query Match 100.0%; Score 19; DB 1; Length 61;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3
 |||
 Db 33 FNP 35

Search completed: August 20, 2002, 11:33:04
 Job time: 1453 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:31:58 ; Search time 191.14 Seconds
(without alignments)
2.715 Million cell updates/sec

Title: US-09-824-286-17

Perfect score: 19
Sequence: 1 FNP 3

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertedrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	11	12	084073
2	19	100.0	16	6	09TOY6
3	19	100.0	18	5	09TWM9
4	19	100.0	18	10	09S817
5	19	100.0	20	2	09R509
6	19	100.0	20	5	025281
7	19	100.0	20	6	P82666
8	19	100.0	21	5	09TWM8
9	19	100.0	22	4	096024
10	19	100.0	22	12	064831
11	19	100.0	23	10	09S821
12	19	100.0	25	2	032326
13	19	100.0	25	11	063988
14	19	100.0	27	5	P90722
15	19	100.0	27	6	09XSC0
16	19	100.0	27	12	068908

17	19	100.0	27	12	091L61	091161	hepatitis c
18	19	100.0	27	12	091L60	091160	hepatitis c
19	19	100.0	27	12	091L59	091159	hepatitis c
20	19	100.0	27	12	091L58	091158	hepatitis c
21	19	100.0	27	12	091L57	091157	hepatitis c
22	19	100.0	27	12	091L56	091156	hepatitis c
23	19	100.0	27	12	091L55	091155	hepatitis c
24	19	100.0	27	12	091L54	091154	hepatitis c
25	19	100.0	27	12	091L53	091153	hepatitis c
26	19	100.0	27	12	091L52	091152	hepatitis c
27	19	100.0	27	12	091L51	091151	hepatitis c
28	19	100.0	27	12	091L50	091150	hepatitis c
29	19	100.0	27	12	091L49	091149	hepatitis c
30	19	100.0	27	12	091L48	091148	hepatitis c
31	19	100.0	27	12	091L47	091147	hepatitis c
32	19	100.0	27	12	091L46	091146	hepatitis c
33	19	100.0	27	12	091L45	091145	hepatitis c
34	19	100.0	27	12	091L44	091144	hepatitis c
35	19	100.0	27	12	091L43	091143	hepatitis c
36	19	100.0	27	12	091L42	091142	hepatitis c
37	19	100.0	27	12	091L41	091141	hepatitis c
38	19	100.0	27	12	091L40	091140	hepatitis c
39	19	100.0	27	12	091L39	091139	hepatitis c
40	19	100.0	27	12	091L38	091138	hepatitis c
41	19	100.0	27	12	091L37	091137	hepatitis c
42	19	100.0	27	12	091L36	091136	hepatitis c
43	19	100.0	27	12	091L35	091135	hepatitis c
44	19	100.0	27	12	091L34	091134	hepatitis c
45	19	100.0	27	12	091L33	091133	hepatitis c

ALIGNMENTS

RESULT 1
ID 084073 PRELIMINARY; PRT; 11 AA.
AC 084073;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 08, Last sequence update)
DE INFLUENZA A/EV/ROSTOCK/34 (H7N1), POLYMERASE 3 (SEG 3), 3' END OF
DE VNNA (INITIATOR REGION FOR PROTEIN CODING) (FRAGMENT).
OS Influenza A virus.
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80034428; PubMed=493121;
RA Robertson J.S.;
RT "5' and 3' terminal nucleotide sequences of the rna genome segments of
RL Influenza virus.",
RL Nucleic Acids Res. 6:3745-3757(1979).
FT EMBL; J02123; AAA43612.1; -.
FT NON_TER 11 11
SQ SEQUENCE 11 AA; 1400 MW; CC2007F7AC412C9 CRC64;

Query Match 100.0%; Score 19; DB 12; Length 11;
Best local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 FNP 3
Db 9 FNP 11

RESULT 2
ID 09TOY6 PRELIMINARY; PRT; 16 AA.
AC 09TOY6;
DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE ANNEXIN I (FRAGMENT).
 OC Oryctolagus cuniculus (Rabbit).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=96226532; PubMed=8967522;
 RA Mayran N., Traverso V., Maroux S., Massey-Harroche D.;
 RT "Cellular and subcellular localizations of annexin I, IV, and VI in
 lung epithelia.";
 RL Am. J. Physiol. 270:L863-L871(1996).
 SO SEQUENCE 16 AA; 1546 MW; 70AA0A45193C8BB6 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 19; DB 6; Length 16;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 Db 14 FNP 16

RESULT 3
 O9TWM9 PRELIMINARY; PRT; 18 AA.
 AC O9TWM9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE NTPAG ANTIGEN (FRAGMENT).
 CC Tetrahymena pyriformis.
 CC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 CC Tetrahymenina; Tetrahymena.
 OX NCBI_TaxID=5908;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=94323706; PubMed=7519359;
 RA Leary J.H.3rd, Evans D.L., Jaso-Friedmann L.;
 RT "Partial amino acid sequence of a novel protozoan parasite antigen
 that inhibits non-specific cytotoxic cell activity.";
 RL Scand. J. Immunol. 40:158-164(1994).
 SO SEQUENCE 18 AA; 2082 MW; 13BCB737B1FF92A3 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 19; DB 5; Length 18;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 Db 7 FNP 9

RESULT 4
 O9S817 PRELIMINARY; PRT; 18 AA.
 AC O9S817;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE 40 KDA pI 8.5 ABSCTISSIC ACID-INDUCED HISTIDINE RICH PROTEIN
 (FRAGMENT).
 CC Oryza sativa (Rice).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE.

RX MEDLINE=95175599; PubMed=7870812;
 RA Moons A., Bauw G., Prinsen E., Van Montagu M., Van der Straeten D.;
 RT "Molecular and physiological responses to abscisic acid and salts in
 roots of salt-sensitive and salt-tolerant indica rice varieties.";
 RL Plant Physiol. 107:177-186(1995).
 SO SEQUENCE 18 AA; 2094 MW; 0CD245DB237E7520 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 19; DB 10; Length 18;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 Db 4 FNP 6

RESULT 5
 O9R509 PRELIMINARY; PRT; 20 AA.
 AC O9R509;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ADENYLATE KINASE (EC 2.7.4.3) (FRAGMENT).
 CC Bacillus alcalophilus.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1445;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=94117390; PubMed=8288548;
 RA Gilles A.M., Glaser P., Perrier V., Meier A., Longin R., Sebald M.,
 RA Maignan L., Pistotnik E., Barzu O.;
 RT "Zinc, a structural component of adenylate kinases from gram-positive
 bacteria.";
 RL J. Bacteriol. 176:520-523(1994).
 DR InterPro: IPR000850; Adenylate_kin.
 DR Pfam: PF00406; adenylatekinase; 1.
 SO SEQUENCE 20 AA; 2190 MW; 94D98FC289220254 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 19; DB 2; Length 20;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 Db 6 FNP 8

RESULT 6
 Q25281 PRELIMINARY; PRT; 20 AA.
 AC Q25281;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ALPHA-TUBULIN DNA (FRAGMENT).
 CC Leishmania enriettii.
 CC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5663;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84142075; PubMed=6321982;
 RA Wirth D.F., Slater C.;
 RT "Isolation and characterization of an alpha-tubulin gene from
 Leishmania enriettii.";
 RL Mol. Biochem. Parasitol. 9:83-92(1983).
 DR EMBL: M28001; AAA29273.1; -
 DR InterPro: IPR003008; Tubulin_FtsZ.
 DR Pfam: PF00091; tubulin; 1.
 KW GTP-binding.

FT NON_TER 1 1
 FT NON_TER 20 20
 SQ SEQUENCE 20 AA; 2351 MW; B9A8F2D1755D810D CRC64;

Query Match 100.0%; Score 19; DB 5; Length 20;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 Db 17 FNP 19

RESULT 7
 ID P82666 PRELIMINARY; PRT; 20 AA.
 AC P82666;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE MITOCHONDRIAL 28S RIBOSOMAL PROTEIN S23 (MRP-S23) (FRAGMENTS).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=LIVER;
 RX MEDLINE=20490686; PubMed=10938081;
 RA Koc E.C., Burkhardt W., Blackburn K., Moseley A., Koc H.,
 RA Spremulli L.L.;
 RT "A proteomics approach to the identification of mammalian
 RT mitochondrial small subunit ribosomal proteins.";
 RL J. Biol. Chem. 275:32585-32591(2000).
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
 KW Ribosomal protein; Mitochondrion.
 FT NON_TER 1 1
 FT NON_CONS 10 11
 FT NON_TER 20 20
 SQ SEQUENCE 20 AA; 2249 MW; EC7FE3CA50071EE4 CRC64;

Query Match 100.0%; Score 19; DB 6; Length 20;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 Db 15 FNP 17

RESULT 8
 ID Q9TWM8 PRELIMINARY; PRT; 21 AA.
 AC Q9TWM8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DE 01-JUN-2000 (TREMBLrel. 14, last annotation update)
 DE NKTAG ANTIGEN (FRAGMENT).
 OS Tetrahymena pyriformis.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 OC Tetrahymenida; Tetrahymena.
 OC NCBI_TaxID=5908;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=94323706; PubMed=7519359;
 RA Leary J.H.3rd, Evans D.L., Jaso-Friedmann L.;
 RT "Partial amino acid sequence of a novel protozoan parasite antigen
 RT that inhibits non-specific cytotoxic cell activity.";
 RL Scand. J. Immunol. 40:158-164(1994).
 SQ SEQUENCE 21 AA; 2353 MW; E0B7ED0F061127B1 CRC64;

Query Match 100.0%; Score 19; DB 5; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 Db 7 FNP 9

RESULT 9
 ID 096024 PRELIMINARY; PRT; 22 AA.
 AC 096024;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE MITOCHONDRIAL RIBOSOMAL PROTEIN S23 (FRAGMENT).
 GN MRPS23.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429115; PubMed=11543634;
 RA Kenmochi N., Suzuki T., Uechi T., Magoori M., Higa S.,
 RA Watanabe K., Tanaka T.;
 RT "The human mitochondrial ribosomal protein genes: Mapping of 54 genes
 RT to the chromosomes and implications for human disorders.";
 RL Genomics 37:65-70(2001).
 DR EMBL: AB061206; BAB54956.1; -
 KW Ribosomal protein.
 FT NON_TER 1 1
 FT NON_TER 22 22
 SQ SEQUENCE 22 AA; 2536 MW; A079B804B5E5E7A6 CRC64;

Query Match 100.0%; Score 19; DB 4; Length 22;
 Best Local Similarity 100.0%; Pred. No. 4.8e+02;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 Db 12 FNP 14

RESULT 10
 ID 064831 PRELIMINARY; PRT; 22 AA.
 AC 064831;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
 DE I5 FIBER (FRAGMENT).
 OS Human adenovirus type 5.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OC NCBI_TaxID=28285;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85092388; PubMed=2981456;
 RA Cladarras C., Mold W.S.;
 RT "DNA sequence of the early E3 transcription unit of adenovirus 5.";
 RL Virology 140:28-43(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85092389; PubMed=3966299;
 RA Cladarras C., Bhat B., Mold W.S.;
 RT "Mapping the 5' ends, 3' ends, and splice sites of mRNAs from the
 RT early E3 transcription unit of adenovirus 5.";
 RL Virology 140:44-54(1985).
 DR EMBL: X03002; CAA26788.1; -

FT NON TER 22 22
SQ SEQUENCE 22 AA: 2575 MW: CA394D2EE31FE71C CRC64;

Query Match
Best Local Similarity 100.0%; Score 19; DB 12; Length 22;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
Db 11 FNP 13

RESULT 11
OY9821 PRELIMINARY; PRT; 23 AA.
AC OY9821;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DS LIGHT-HARVESTING ANTENNA OF PHOTOSYSTEM I 680A, LHCI 680A.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_Taxid=4513;
RN [1]
RP SEQUENCE.
RX MEDLINE=92267013; PubMed=1587270;
RA Kneetzel J., Svendsen I., Simpson D.J.;
RT "Identification of the photosystem I antenna polypeptides in barley.
RT Isolation of three pigment-binding antenna complexes.";
RL Eur. J. Biochem. 206:209-215(1992).
SQ SEQUENCE 23 AA: 2282 MW: 194B21969A17581E CRC64;

Query Match
Best Local Similarity 100.0%; Score 19; DB 10; Length 23;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
Db 17 FNP 19

RESULT 12
OY32326 PRELIMINARY; PRT; 25 AA.
AC OY32326;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE REGION UPSTREAM OF THE KDPDE OPERON, KDP A, KDP B AND KDP C GENES.
OS Clostridia; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_Taxid=1488;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-DSM 792;
RX MEDLINE=97369818; PubMed=9226259;
RA Treuner-Lange A., Kuhn A., Durre P.;
RT "The kdp system of Clostridium acetobutylicum: cloning, sequencing,
RT and transcriptional regulation in response to potassium
RT concentration.";
RL J. Bacteriol. 179:4501-4512(1997).
DR EMBL: U44892; AAC45476.1; -
SQ SEQUENCE 25 AA: 3023 MW: 8D64226018BA2254 CRC64;

Query Match
Best Local Similarity 100.0%; Score 19; DB 2; Length 25;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
Db 20 FNP 22

RESULT 13
OY63988 PRELIMINARY; PRT; 25 AA.
AC OY63988;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HOX2.4 HOMEOBOX HOMOLOG PROTEIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-W1STAR.
RC TISSUE-BONE MORPHOGENETIC PROTEIN-IMPLANTED SUBCUTANEOUS MUSCLE;
RX MEDLINE=94271262; PubMed=7911662;
RA Iimura T., Oida S., Takeda K., Maruoka Y., Sasaki S.;
RT "Changes in homeobox-containing gene expression during ectopic bone
RT formation induced by bone morphogenetic protein.";
RL Biochem. Biophys. Res. Commun. 201:980-987(1994).
DR EMBL: S71284; AAB31004.2; -
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
KW Homeobox; DNA-binding; Nuclear protein.
FT NON_TER 1 1
FT NON_TER 25 25
SQ SEQUENCE 25 AA: 3111 MW: 5B88C1D41ACB974 CRC64;

Query Match
Best Local Similarity 100.0%; Score 19; DB 11; Length 25;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
Db 2 FNP 4

RESULT 14
OY90722 PRELIMINARY; PRT; 27 AA.
AC OY90722;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HOMEOBOX PROTEIN BHOX 43 (FRAGMENT).
OS Beroe ovata.
OC Eukaryota; Metazoa; Ctenophora; Cycliocoela; Beroidea; Beroe.
OX NCBI_Taxid=10201;
RN [1]
RP SEQUENCE FROM N.A.
RA Witelch H.J., Meech R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U83686; AAB41253.1; -
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PRINTS: PR00024; HOMEOBOX.
FT NON_TER 1 1
FT NON_TER 27 27
SQ SEQUENCE 27 AA: 3269 MW: 9861CB88C70F14AC CRC64;

Query Match
Best Local Similarity 100.0%; Score 19; DB 5; Length 27;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
 |||
 Db 2 FNP 4

RESULT 15

O9XSCO PRELIMINARY; PRT; 27 AA.
 ID O9XSCO
 AC O9XSCO;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE BETA-LACTOGLOBULIN C (FRAGMENT).
 GN LGB.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OX NCBI_TaxID=9940;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BREED MERINO-LAND;
 RA Prinzenberg E.-M., Erhardt G.;
 RT "Molecular genetic characterization of ovine beta-lactoglobulin C
 precursor allele and detection by PCR-RFLP.";
 RL J. Anim. Breed. Genet. 0:0-0(1999).
 DR EMBL; AF117693; AAD30423.1; -.
 DR HSSP; P02754; 1CJ5. 1
 FT NON_TER 1
 SO SEQUENCE 27 AA; 3051 MW; A4AC05426D1DA34 CRC64;

Query Match 100.0%; Score 19; DB 6; Length 27;
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 FNP 3
 |||
 Db 16 FNP 18

Search completed: August 20, 2002, 11:31:59
 Job time: 1473 sec

THIS PAGE BLANK (USPTO)